On the Prosody of Orkney and Shetland Dialects

Leyden, K. van; Heuven, V.J. van

Original Paper

Phonetica 2006;63:149–174
DOI: 10.1159/000095306

Received: February 13, 2006
Accepted: June 1, 2006

On the Prosody of Orkney and Shetland Dialects

Klaske van Leyden, Vincent J. van Heuven

Phonetics Laboratory, Leiden University Centre for Linguistics (LUCL), Leiden, The Netherlands

Abstract

The aim of this study is to find experimental support for impressionistic claims that there are prosodic differences between the dialects of Orkney and Shetland. It was found that native listeners had no difficulty in discriminating between Orkney and Shetland dialects when presented with speech fragments containing only melodic information. The results of a subsequent acoustic investigation revealed that there is a striking difference in pitch-peak location, which can be characterised as a shift in the location of the entire rise, i.e. both the onset and the peak. Shetland has early alignment, whereas the accent-lending rise in Orkney occurs late, so that in disyllabic words with initial stress the pitch peak does not coincide with the stressed syllable, but is delayed until the post-stress syllable. Finally, the perceptual relevance of the prosodic parameters identified in the acoustic study was investigated.

1 Introduction

Orkney and Shetland – groups of islands to the north of mainland Scotland – were colonised by Vikings from Norway in the 9th century.[1] The Scandinavian language spoken by the settlers and their descendants became known as Norn and remained the chief medium of communication for several centuries. After the islands became part of Scotland in the second half of the 15th century, Scottish administrators and landowners, bringing with them the Scots language, began to arrive in Orkney and Shetland. Through a process of language shift, Norn gradually became restricted to the family circle, until in the course of the 18th century it was replaced by Lowland Scots.

The dialects currently spoken in Orkney and Shetland are conservative varieties of Lowland Scots with a substantial Scandinavian substratum. However, Shetland appears

[1]Preliminary versions of experiments 3, 4 and 5 appeared as van Leyden and van Heuven [2003], van Heuven and van Leyden [2003] and van Leyden [2004]. Section 5 (experiment 4) is a revised version of van Leyden [2006]; the authors are grateful to Gösta Bruce for granting permission to reuse parts of this paper.

KARGER

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

0031–8388/06/0633–0149
$23.50/0
Accessible online at:
www.karger.com/journals/pho

Klaske van Leyden
Phonetics Laboratory, Universiteit Leiden
PO Box 9515
NL–2300 RA Leiden (The Netherlands)
Tel. +31 71 527 2319, Fax +31 71 527 2615
E-Mail kvleyden@wxs.nl

to have retained its Scandinavian substratum to a greater degree than has Orkney [van Leyden, 2002, 2004]. Dialect differences are apparent not only at lexical and syntactic levels but also with respect to segmental phonology [Johnston, 1997] and the temporal organisation of the syllable [Catford, 1957; van Leyden, 2002]. The main difference between the two varieties, however, appears to be a dissimilarity in intonation. Impressionistically, Shetland speech has a somewhat narrow pitch range and, with respect to intonation, Shetland dialect seems not dissimilar from most other dialects of Scots, while Orcadian stands out, being characterised by a very distinctive, rise-fall intonation pattern, referred to as a lilting or sing-song melody by the local population.

Although the natives themselves are conscious of the melodic differences between the two dialects, there appear to be no published works dealing with the intonation of Orkney and Shetland speech. The aim of the present study, therefore, is to investigate the validity of impressionistic claims that there are prosodic differences between the dialects of Orkney and Shetland. To begin with, we will examine the role of intonation versus segmental information in the identification of Orkney and Shetland dialects by native listeners of each of these varieties (experiments 1 and 2). We will then carry out an acoustic investigation of the prosodic parameters of the two dialects (experiment 3). Finally, the relevance of the established prosodic parameters will be evaluated perceptually (experiment 4).

2 Experiment 1: The Role of Intonation

Native Shetlanders typically claim that it is very easy to identify an Orcadian by his intonation alone. Results of experiments investigating the role of intonation in the recognition of dialects indicate that intonation may indeed be a cue in the identification of regional varieties, particularly when identifying one's own dialect [Gooskens, 1997; Schaeffler and Summers, 1999; Gooskens and van Bezooijen, 2002; Peters et al., 2003]. Using an experimental design similar to that of Gooskens and van Bezooijen [2002], we therefore presented speech fragments both with and without the original intonation contour, in order to examine the contribution of prosody (more specifically, intonation) in the identification of Orkney and Shetland dialects.

2.1 Method

Short fragments of spontaneous speech were presented to native listeners of Orkney and Shetland. The fragments were taken from brief, informal interviews recorded during earlier fieldwork trips by the first author. Six male speakers were selected; 3 from Orkney (West Mainland, Kirkwall and Westray) and 3 from Shetland (Burra Isle, West Mainland and North Mainland). Female speakers were not included, because applying the same low-pass filter for both male and female voices leaves relatively more traces of spectral information in the female speech. Hence, inclusion of both genders could possibly introduce an additional variable that would have interfered with our results. The sole selection criterion was that the selected speakers had to come from different parishes throughout Orkney and Shetland and had produced ample fluent speech during the interview; none of the speakers spoke very broad dialect. All speakers were between 35 and 50 years of age. The selected speech fragments were about 12 s in duration and comprised one or more syntactic sentences of semantically neutral content; only one fragment per speaker was included.

Two speech conditions were created: (1) normal (intelligible) speech and (2) LP-filtered (unintelligible) speech. LP-filtering delexicalises the speech signal by removing most of the spectral information while leaving the temporal organisation (intonation and syllable structure) intact. In this way, it is possible to investigate how well listeners are able to identify a language variety on the basis of

prosody only. Using Praat speech processing software [Boersma and Weenink, 2005], the speech signal was LP-filtered at 300 Hz, with a band smoothing of 50 Hz.

In both normal and filtered speech, two intonation conditions were generated: (1) original speech melody and (2) monotonised speech. By eliminating the intonation contour (through monotonisation), we are able to establish the importance of speech melody in distinguishing between the two dialects at issue. The condition with unintelligible, monotonous speech allows us to examine the role of temporal organisation and intensity in the identification of Orkney and Shetland dialects. (Note that Gooskens and van Bezooijen [2002] omitted this condition. Note also that in previous studies [e.g. Cohen and 't Hart, 1970] listeners were apparently unable to distinguish between language varieties when presented with unintelligible, monotonous speech.) Monotonisation was done with Praat software by changing the pitch contour into a flat line, using PSOLA analysis and resynthesis [Moulines and Verhelst, 1995]. This line was given no declination, since the distinguishing role of declination in Orkney and Shetland dialects is not known.

The manipulated speech fragments were converted from digital (16 kHz, 16 bit) to analog (allowing a signal bandwidth of just less than 8 kHz) and then recorded onto minidisc using a Sony MZ-R35 minidisc recorder and organised into four blocks, with fragments randomised within each block: (1) LP-filtered speech, monotonised; (2) LP-filtered speech, original intonation contour; (3) original speech, monotonised and (4) original speech, original intonation contour. The fragments were the same in each condition (block).

In order to compensate for a possible learning effect, two counterbalanced lists were created for each block. The interstimulus interval was 4 s (offset to onset); an alert tone was played 1 s prior to each stimulus onset.

2.2 Subjects and Procedure

Twenty listeners took part in the experiment, 10 from Orkney (6 male and 4 female) and 10 from Shetland (6 male and 4 female). The subjects were chosen from several parishes throughout Orkney and Shetland and were between 30 and 50 years of age, having resided locally most of their lives. The listeners reported no hearing problems and were not paid for their participation.

The subjects were tested individually in their own home or workplace in experimental sessions lasting about 15 min including instruction time. The stimuli were presented over headphones (Sennheiser HD 455) at a comfortable listening level. Subjects were issued with response sheets on which they were asked to indicate, for each utterance, where they thought a particular speaker hailed from. They responded by ticking on a 10-point scale running between 1 'definitely from elsewhere' and 10 'definitely from Orkney' (for Orkney listeners) or 'definitely from Shetland' (for listeners in Shetland). Subjects were required always to tick a scale position, even if they felt they had to guess. It should be noted that, since the scale has no midpoint, the listeners were always forced to make a decision, no matter how tentative.

The four blocks of stimuli were played in the following order: (1) LP-filtered speech, monotonised; (2) LP-filtered speech, original intonation contour; (3) original speech, monotonised; (4) original speech, original intonation contour. That is, the presentation of material was so arranged that the amount of linguistic information that was made available to the listeners increased from one block to the next, so as to prevent listeners from transferring information gathered from one block over to the next. Each block was preceded by two practice fragments, recorded by speakers other than those used in the experiment; responses to these trials were not included in the analysis.

2.3 Results

A total of 20 (subjects) $\times$ 6 (speakers) $\times$ 4 (presentation conditions) $-$ 1 (missing response) $=$ 479 responses were collected. The judgement scores were analysed by separate two-way analyses of variance (ANOVA) for intelligible and unintelligible speech, further broken down by listener group (Orkney and Shetland), assuming fixed factors for speech condition and dialect (Orkney or Shetland). Figure 1 presents the mean judgement scores broken down by presentation condition and by dialect.

As can be seen in figure 1, for Orkney listeners, there is a large effect resulting from the dialect of the speaker, $F(1, 231) = 313.8$ ($p < 0.001$), which strongly interacts with

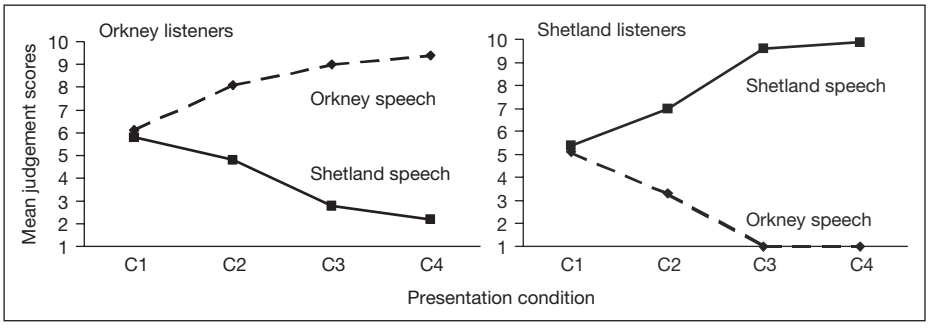

Fig. 1. Spontaneous speech: mean judgement scores (1 'definitely elsewhere'; 10 'definitely my island'), broken down by presentation condition (C1–4) and by dialect, for Orkney listeners (left) and Shetland listeners (right). C1 = LP-filtered and monotonous; C2 = LP-filtered and original intonation contour; C3 = normal speech and monotonous; C4 = normal speech and original intonation contour. Mean data of 10 subjects per listener group.

presentation condition, $F(3, 231) = 42.8$ ($p < 0.001$), with judgement scores differentiating incrementally as a function of the linguistic cues that were made available to the listener. Similar, in fact even larger effects are observed for Shetland listeners, $F(1, 232) = 919.7$ ($p < 0.001$) for the effect of speaker dialect and $F(3, 232) = 134.3$ ($p < 0.001$) for the dialect × condition interaction.

Clearly, condition C1 (monotonous, unintelligible speech) contains no audible information that allows our listeners to differentiate between the two varieties. When the intonation contour is preserved in the unintelligible speech condition (C2), the origin of the speakers is distinguished rather well. However, when monotonised, *intelligible* speech is presented (C3), the differentiation between the varieties is nearly maximal. From this outcome it can be concluded that the two dialects differ distinctly with respect to their segmental structure. Finally, combining both information sources (C4) yields even better differentiation.

2.4 Conclusions

The results of experiment 1 show that native listeners from Orkney and Shetland distinguish quite clearly between the two intonational systems when they are presented with unintelligible speech samples in both dialects, i.e. when only prosodic (melodic and temporal) information is available. However, the two varieties were indistinguishable when listeners heard speech that was both monotonised and unintelligible. Therefore, the conclusion is warranted that the prosodic difference is a matter of intonation rather than temporal organisation. Furthermore, it was shown that the two dialects also differ distinctly with respect to their segmental structure.

3 Experiment 2: Intonation versus Segments

Although the elimination technique as used in experiment 1 is suitable for demonstrating the role of intonation in distinguishing between dialects, it does not allow us to quantify the relative importance of intonation and segmental information in the perceptual

identification of the language varieties. For example, even though we found that adding segmental information to condition C1 affords better identification of dialect than adding intonation, there is no guarantee that segmental information overrides intonation in a direct manner. In experiment 2, therefore, we examined the relative contribution of intonation and segmental information to the mutual identifiability of the target dialects by presenting speech both with and without the original melody, as well as speech fragments in which segmental and prosodic information were in conflict, i.e. by artificially exchanging the pitch curves between realizations of the same sentence in the two dialects while keeping the segmental information unaffected. Obviously, if an utterance spoken by a Shetlander (i.e. containing Shetland segmental information) but with an Orcadian intonation pattern is identified as produced by a speaker from Orkney, then intonation is the stronger of the two sets of cues. If, on the other hand, the listeners judge the hybrid utterance to be spoken by a Shetlander, then the segmental properties outweigh the intonational cues.

3.1 Method

It can be expected that spontaneous speech is most varied with respect to intonation [cf. Cruttenden, 1997, p. 128]; however, for the investigation at hand, it was decided to use speech that was read aloud, rather than spontaneous speech, as it would have been too difficult to elicit suitable speech material during interviews.

Recordings of the utterance *There are many gardens in Bergen*, pronounced with a pitch accent on *gardens*, were selected from the corpus collected for the acoustic study (experiment 3). Four male speakers were selected: Orkney 1 (or1), Orkney 2 (or2), Shetland 1 (sh1) and Shetland 2 (sh2).[2] The main criteria in the selection of the speakers were matching voice quality as well as speaking rate. The Orkney speakers both originated from Kirkwall, the Shetlanders were from West Mainland; all were aged about 40.

Two speech conditions were generated: (1) LPC-resynthesised speech (intelligible), (2) 'buzz' (unintelligible).[3] Using Praat software, a buzz was created by replacing segmental information by a spectrally invariant buzz-like sound (sawtooth wave) while preserving the amplitude and fundamental frequency (F_0) modulations of the original. Consequently, all spectral information was removed from the speech signal but prosodic variation was maintained. In this way, we could be certain that the perceptual identification of a particular language variety was based on prosodic cues only. (It may be objected that LP filtering or replacing spectral information by a spectrally invariant buzz obscures the boundaries between successive segments. Since the alignment of tonal targets may depend on the precise location of certain segment boundaries, an important characteristic of the intonation pattern may get lost. Our results show, nevertheless, that our listeners clearly distinguished between alignments, presumably because the overall intensity contours of syllables were unaffected.) LPC resynthesis (autocorrelation method; Markel and Gray [1976]) was carried out in order to conceal the identity of the speakers.[4] (In these small island communities, there is a good chance that listeners personally know

[2]A Standard Scottish English speaker from mainland Scotland was included as a control condition. However, for the sake of clarity the responses to these stimuli will not be analysed in the present paper; for further details see van Leyden [2004].

[3]In earlier unpublished reports of this research [van Leyden and van Heuven, 2003; van Leyden, 2004] a second unintelligible condition was used, i.e. low-pass-filtered speech (as was also employed in experiment 1 in the present article). Since, theoretically, information may be left in the filtered signal that may allow listeners to distinguish the segments, we preferred buzzed speech to LP-filtered speech. The results in our unpublished studies, however, show that the results obtained for the two types of unintelligible speech did not significantly differ from each other so that the choice is immaterial here.

[4]In LPC resynthesis as applied here, i.e. without adding the residual error signal to the output, the voice quality of the speaker is affected to some extent, as the excitation signal is no longer the speaker's original glottal pulse but an artificial sawtooth wave. Because of this substitution, LPC-resynthesised voices should resemble each other more than the originals. Still, speakers remain more or less identifiable since the resonance characteristics of the supraglottal tract (i.e. the formant structure) of the original speaker are maintained in the resynthesis.

Table 1. Overview of stimulus types

Speaker	Pitch contour				
	or1	or2	sh1	sh2	monotonous
OR1	+	+	+	+	+
OR2	+	+	+	+	+
SH1	+	+	+	+	+
SH2	+	+	+	+	+

See text for explanation of labels.

one or more of the speakers.) Speech was analysed and resynthesised with five formants and the associated bandwidths in the 0- to 5-kHz range, with a sawtooth wave as the source signal in the resynthesis; crucially, no residue (or error signal) was used in the resynthesis. The resulting quality was highly intelligible but – at least in the short utterances used for our experiment – masked the speaker's identity to some extent. (Listeners were debriefed after their participation in the experiment. In no case could a participant identify any of the speakers they had heard.)

For each of the two speech conditions (buzz and LPC-resynthesised), three types of test utterance were generated: (1) monotonous; (2) original intonation (stylised); (3) transplanted melodies. Monotonisation was done by changing the pitch contour into a flat line; this line was given no declination (cf. experiment 1). The utterances were monotonised at 100 Hz.

Stylised versions of the original intonation contours were made in order to afford easy exchange of melodies between segmentally similar utterances (condition 3 'transplanted melodies'). The stylisations were so-called close copies as defined within the IPO tradition of intonation research ['t Hart et al., 1990]. A close copy is defined as 'a synthetic approximation of the natural course of pitch, meeting two criteria: it should be perceptually indistinguishable from the original, and it should contain the smallest possible number of straight-line segments with which this perceptual equally can be achieved' [Nooteboom, 1997, p. 646]. In the IPO tradition, the straight lines are defined in logarithmic plots (F_0 in semitones) as a function of linear time.

The following transplantations were implemented: Orkney contours grafted onto the Shetland utterances (SH1 and SH2) and Shetland contours superimposed on the Orkney utterances (OR1 and OR2). Furthermore, to make the design fully orthogonal, the contours of the 2 Orkney speakers (or1 and or2) were interchanged, as were the two Shetland contours (sh1 and sh2). For each original utterance, F_0 was extracted (autocorrelation method) and interactively stylised, allowing at most one linear rise and one linear fall per syllable. The time coordinates of the pivot points in the resulting rise-fall sequence were expressed relative to the onset and offset of the syllable. The same relative timing of rises and falls was observed after transplantation of the contour; the frequency values of the transplanted contours were left as measured in the original environment. The four intonation contours are illustrated in the 'Appendix'; table 1 gives an overview of the stimulus types. (In the designation of the manipulations, the first part of the labels, in capitals, refers to the origin of the segmental information; the second part, in lower case, refers to the origin of the pitch contour.)

The manipulated utterances were converted from digital (16 kHz, 16 bit) to analog (allowing a signal bandwidth of just less than 8 kHz) and then recorded onto minidisc using a Sony MZ-R35 minidisc recorder and organised into four blocks as follows: (1) buzz and monotonised; (2) buzz and (manipulated) intonation contour; (3) LPC-resynthesised and monotonised; (4) LPC-resynthesised and (manipulated) intonation contour. The stimuli were randomised within each block. Two counterbalanced lists were created for each block, so as to cancel possible learning effects.

Because of the shortness of the stimuli (about 1.5 s), it was decided to present each test utterance twice in succession, with an interval of 1 s separating the two tokens. The interstimulus interval was 3.5 s (offset of second token to the onset of the first token of the next stimulus); an alert tone was played 1 s prior to each stimulus onset.

Table 2. Grouping of responses

Speaker	Pitch contour				
	or1	or2	sh1	sh2	monotonous
OR1 OR2	ORor		ORsh		OR
SH1 SH2	SHor		SHsh		SH

See text (section 3.1) for explanation of labels.

3.2 Subjects and Procedure

Thirty-nine listeners took part in the experiment, 19 from Orkney (10 male and 9 female) and 20 from Shetland (10 male and 10 female). The subjects were between 30 and 50 years of age. They reported no hearing problems and were not paid for their participation. Some of the listeners had also participated in experiment 1. The experimental procedure was identical to the procedure described in section 2.2 above.

The four blocks of stimuli were played in the following order: (1) buzz and monotonised; (2) buzz and (manipulated) intonation contour; (3) LPC-resynthesised and monotonised; (4) LPC-resynthesised and (manipulated) intonation contour (cf. experiment 1). Each block was preceded by two practice fragments; responses to these trials were not included in the analysis. As in experiment 1, the listeners had to respond by ticking on a 10-point scale running between 1 'definitively from elsewhere' and 10 'definitively my island'.

3.3 Results

The experiment yielded a total of 39 (subjects) $\times$ 20 (manipulations) $\times$ 2 (speech conditions: buzz and LPC-resynthesised) = 1,560 responses.

As can be seen in table 1 above, there are five types of pitch contour and 4 speakers, however, to make our dataset manageable, it was decided to group the responses together as listed in table 2.

Figure 2 presents the results of experiment 2 in four panels. Figure 2a and c refer to the results obtained for unintelligible (buzzed) speech and figure 2b and d to intelligible (LPC) speech. The upper panels (fig. 2a, b) are based on the judgements given by Orkney listeners whilst the bottom panels (fig. 2c, d) display the responses for Shetland listeners. In each panel the judgement scores are plotted as a function of the type of melody (Orkney, monotonous, Shetland) and broken down further by segmental information (Orkney versus Shetland). The data were submitted to a Repeated Measures ANOVA (RM-ANOVA) with melody, segments and speech condition (buzz, LPC) as within-subject factors, and with the dialect of the listener (Orkney, Shetland) as a between-subjects factor. Degrees of freedom are always Huynh-Feldt-corrected. The results of the ANOVA are summarised in table 3.

There are two significant main effects, i.e. of melody and of origin of the speaker's segments. Overall, the judgement scores are close to the midpoint of the scale (5.5), with 5.3 for Orkney melody, 5.7 for Shetland and 5.8 for no melody at all (monotonous). Similarly, stimuli with Orkney segments (across all conditions) were judged at 5.3 against 5.8 for Shetland segments. These main effects are irrelevant to our purpose, and merely indicate a bias on the part of Shetland listeners for rejecting Orkney speech stronger than for Orkney listeners to reject Shetland speech (whether segmentally or

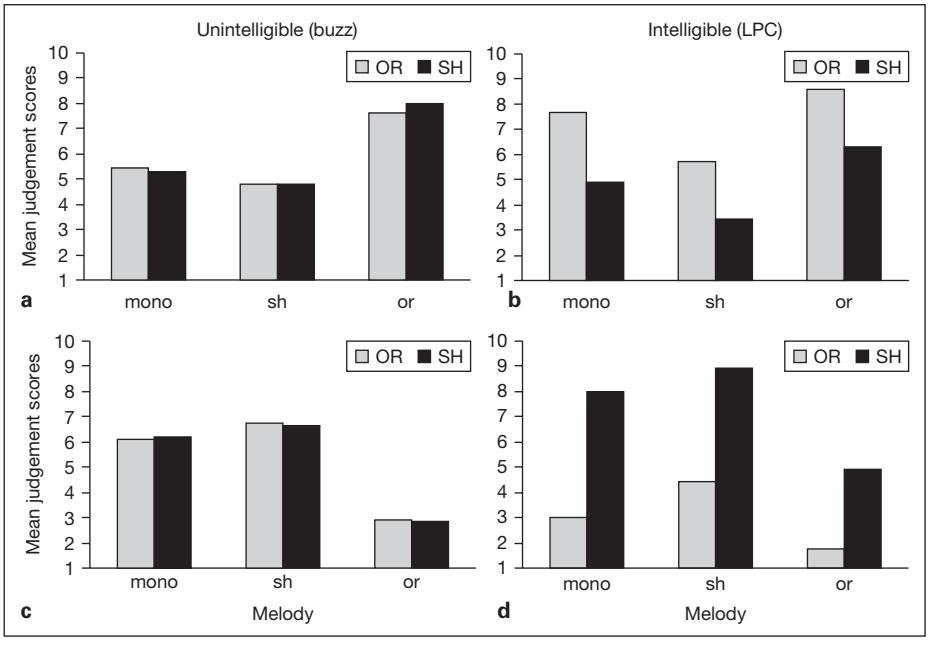

Fig. 2. Mean judgement scores (1 'definitely elsewhere'; 10 'definitely my island') for unintelligible speech (**a**, **c**) and intelligible speech (**b**, **d**); Orkney listeners top (**a**, **b**), Shetland listeners bottom (**c**, **d**). The scores are broken down by melody (monotonous, or, sh) and segmental information (OR, SH). See text for explanation of labels. Mean data of 19 Orkney and 20 Shetland subjects.

Table 3. Summary of RM-ANOVA for experiment 2

Effect/interaction	Df_1	Df_2	**F**	p
Speech condition (C)	1	37	0.070	=0.792
Segments (S)	1	37	13.2	=0.001*
Melody (M)	2	62.9	4.7	=0.017*
Listener group (L)	1	37	25.0	<0.001*
C × S	1	37	7.2	=0.011*
C × M	2	74	0.599	=0.552
C × L	1	37	0.262	=0.612
S × M	2	74	2.3	=0.104
S × L	1	37	174.7	<0.001*
M × L	2	62.9	209.8	<0.001*
C × S × M	2	71.7	1.1	=0.331
C × S × L	1	37	103.0	<0.001*
C × M × L	2	74	12.0	<0.001*
S × M × L	2	74	1.7	=0.183
C × S × M × L	2	71.7	5.7	=0.005*

melodically cued). This bias is even more apparent when we break down the judgement scores by listener group (6.1 for Orkney listeners against Shetland listeners 5.1). However, we are primarily interested in the interactions. Three second-order interactions prove significant. Obviously, listeners react very differently to speech with segments suggesting their own dialect than they do to segments of the other dialect – a difference which is found only for intelligible speech, so that the third-order interaction segments × listeners × speech condition is also highly significant. The interaction between melody and listener group is even stronger. Finally, there is a small interaction between speech condition and segments.

Figure 2d indicates that Shetland listeners clearly differentiate between Shetland segmental information with original pitch contours (8.9) and monotonised Shetland speech (8.0) on the one hand, and all other pitch manipulations on the other (<5.0). For Orkney listeners (fig. 2b), the judgement scores are more tightly clustered. Speech with Orkney pitch contours, but with Shetland segmental structure, is also classified as Orcadian (i.e. ≥ 6.0). Note that both Orkney and Shetland listeners always judge monotonous stimuli to be 'more native' than stimuli with an intonation contour from the other island group; this is true for both unintelligible and intelligible speech.

When in original Orkney speech (8.6) the pitch contour is replaced by its Shetland counterpart, acceptability scores – as judged by native Orcadians – drop by 2.9 points; if the original intonation is maintained but the segments are replaced by their Shetland counterparts, the scores drop by 2.3 points. Hence, the detrimental effect of replacing the intonation contour is larger (by 0.6 point) than that of replacing the segments. In the complementary situation, acceptability scores as judged by Shetland listeners drop by as much as 4.0 points (from 8.9 to 4.9) when the original Shetland intonation is replaced by the Orkney pattern. The scores drop to even lower values (to 4.4), however, when the original Shetland intonation is preserved and the segments are replaced by their Orkney counterparts (i.e. a drop of 4.5 points).

Thus the results of experiment 2 reveal an asymmetrical effect. The Orkney listeners seem to attach more weight to intonation than to segments, whilst the reverse seems to be the case for the Shetland listeners. Furthermore, the Shetland listeners seem to react more negatively to Orkney speech features, whether prosodic or segmental, than do Orcadians to Shetland speech.

3.4 Conclusions

We aimed to determine the relative contribution of segmental information and intonation contour by artificially creating a conflict between the two information sources. The results of experiment 2 bear out that the contribution of segments and intonation to the acceptability of a speech sample are roughly equal. For Shetland listeners, segmental deviations contribute more to non-nativeness than does a deviant intonation pattern, an effect that has commonly been reported for this type of study. However, for Orcadians, intonation was the stronger cue for non-nativeness.

4. Experiment 3: Acoustic Investigation

As we saw in the previous sections, native listeners proved quite able to distinguish Orkney from Shetland speech on the basis of melodic information only. We now present an acoustic investigation of the melodic and temporal differences between Orkney

and Shetland dialects. A systematic comparison of the acoustic measurements with the perceptual results may allow us to isolate potential cues in the melodic systems by which native listeners differentiate between the two dialects.

4.1 Method

Four short sentences were selected for the production experiment. Two declaratives *There are many gardens in Bergen* and *There are many houses in Bergen* as well as two yes-no questions *Are there many gardens in Bergen*? and *Are there many houses in Bergen*? This type of sentence was chosen in order to permit a comparison with similar investigations, e.g. Thorsen [1978]. Furthermore, since no F_0 values can be extracted from unvoiced segments, the sentences were devised so as to contain only voiced consonants (with the exception of the /h/ of 'houses').

4.2 Subjects and Procedure

Twelve male and 7 female speakers from Orkney and 11 male and 8 female speakers from Shetland (i.e. 38 speakers all together) were recorded for the experiment. The subjects were chosen from several parishes throughout Orkney and Shetland and were between 30 and 50 years of age. They reported no speech or hearing problems and were not paid for their participation.

The four target sentences were recorded together with fourteen other utterances of the type *There are/Are there many… in…* The stimuli were printed individually on cue cards and presented one at a time; the random presentation order differed per speaker. Subjects were instructed to read at a natural speaking rate and to pronounce the sentences with emphasis on *gardens*/*houses* (these words were printed in bold on the cue cards). The speakers were individually recorded onto minidisc (Sony MZ-R35) in their own home or workplace, each speaker recording the list of materials twice, with a short break between the lists. The analog output of the minidisc recordings was AD converted (16 kHz, 16 bit) and stored on computer disk for later analysis.

4.3 Analysis and Results

The data set nominally comprised in total 304 recorded utterances: 38 (speakers) $\times$ 4 (stimuli) $\times$ 2 (repetitions). However, for each of the subjects, only the second recording of the material was acoustically analysed, and only when a speaker mispronounced a particular sentence, which happened in about five tokens, was the first recording used. In addition, the recordings of 1 male Orcadian were discarded because of excessive creak. Thus, the actual data set comprised 37 (speakers) $\times$ 4 (stimuli) = 148 tokens.

Using Praat speech-processing software, F_0 was extracted (autocorrelation method) for each utterance. Rise-fall configurations of the pitch accent on *gardens*/*houses* as well as the accent on *many* were interactively stylised by replacing the original F_0 contour by a perceptually equivalent sequence of straight-line interpolations between selected pivot points, as in van Heuven and Haan [2000], using PSOLA analysis and resynthesis [Moulines and Verhelst, 1995]. It should be noted that, in many instances, the Orkney speakers also had a pitch peak on *Bergen* (most likely a boundary tone), while this occurred only occasionally in the case of the Shetland speakers. It was therefore decided not to analyse the rise on *Bergen*. Without a single exception, however, all speakers had produced a large pitch movement on the word *many* and we assume that all tonal configurations on *many* and *gardens/houses* were accent-lending. The time-frequency coordinates (in milliseconds and Hertz, respectively) of the pivot points defining the pitch movements were stored together with the onsets of the syllables and the duration of the syllables with which the pitch movements were aligned. No frequency values were entered in those cases where the expected rise-fall accent was omitted from a particular constituent by the speaker. This happened in about twelve instances, each time in the case of *many*.

4.3.1 Intonation

The measured frequency values were first converted from Hertz to equivalent rectangular bandwidth [ERB; formula (F in Hz): ERB = 16.6 × log (1 + F/165.4)]. The ERB scale expresses the relationship between pitch values and human perception and is regarded as the most appropriate scale for intonation studies [Hermes and van Gestel, 1991]. This scale also allows a direct comparison of male and female speech.

The test words *many*, *gardens* and *houses* differ considerably with respect to their segmental make-up and first-syllable duration. Therefore, in order to permit comparison of pitch rise location across the three words, it was decided to normalise the alignment of the rises on *many* (R1) and on *gardens/houses* (R2) as follows. The onset of the stressed syllable was given the value of 0% relative time, whilst the offset of the stressed syllable was set at 100%. The onset and offset locations of the pitch rise were then expressed in terms of this relative time scale. Thus, a rise onset at 50% marks a rise the onset of which occurs at a point in time located halfway along the stressed syllable. Similarly, a rise offset at 200% relative time is located two syllable lengths after the onset of the stressed syllable (i.e. a peak located at the syllable following the stress).

Figure 3 presents the basic intonation patterns that were realised on the accents on *many* and *gardens*/*houses*, with pitch plotted in equivalent rectangular bandwidth as a function of time, in separate panels for statements (fig. 3a) and questions (fig. 3b). The first zero point on the time scale coincides with the onset of the first (i.e. stressed) syllable of *many*; while the second zero point coincides with the onset of the first syllable of *gardens*/*houses*. In each panel, the intonation patterns of the male speakers are located in the bottom half of the pitch range, typically between 3 and 5 ERB, whilst the female contours are in the upper half of the range, typically between 5 and 7 ERB. Only the rise portions of the accents were plotted. The dotted lines connecting the end of the rise on *many* to the onset of the rise on *gardens*/*houses* ignore the – highly variable – location of the endpoint of the pitch fall somewhere between the two rises.

The first notable difference between the Orkney and Shetland patterns shown in figure 3 is that, on the whole, the overall pitch in Orkney is substantially higher than in Shetland. Given the fairly large number of speakers involved in this study, it seems safe to rule out the possibility that the observed pitch differences are accidental. When individual distributions of F_0 are compared, it is clear that (allowing for differences in pitch between the sexes) there is tight clustering within each dialect community. Also, the fact that the lower pitch in Shetland recurs both in the male and in the female group indicates that the difference is not likely to be due to infelicitous speaker sampling. Rather, we would surmise that the overall higher pitch in Orcadian is a feature of the dialect.

Secondly, we observe that, both in Orkney and in Shetland, the pitch rises have the same slopes and excursion sizes, at least when expressed in equivalent rectangular bandwidth, for men as they have for women. Therefore, it seems that the only difference that is related to gender is the overall (i.e. mean) pitch.[5]

Thirdly, in statements the pitch rises are roughly equally large (excursion sizes between 0.8 and 0.9 ERB) for *many* and for *gardens*/*houses*, irrespective either of the

[5]The (perceptual) pitch range of male versus female speech is a matter of considerable debate. Although it has repeatedly been claimed that females have a systematically wider range than males, recent studies [e.g. Biemans, 1998, for Dutch, and Henton, 1998, for English] failed to find gender-related differences in pitch range. Yet, Haan [2002] found that, at least for Dutch interrogative sentences, there are differences in excursion size between the genders, with women producing larger excursion sizes than men.

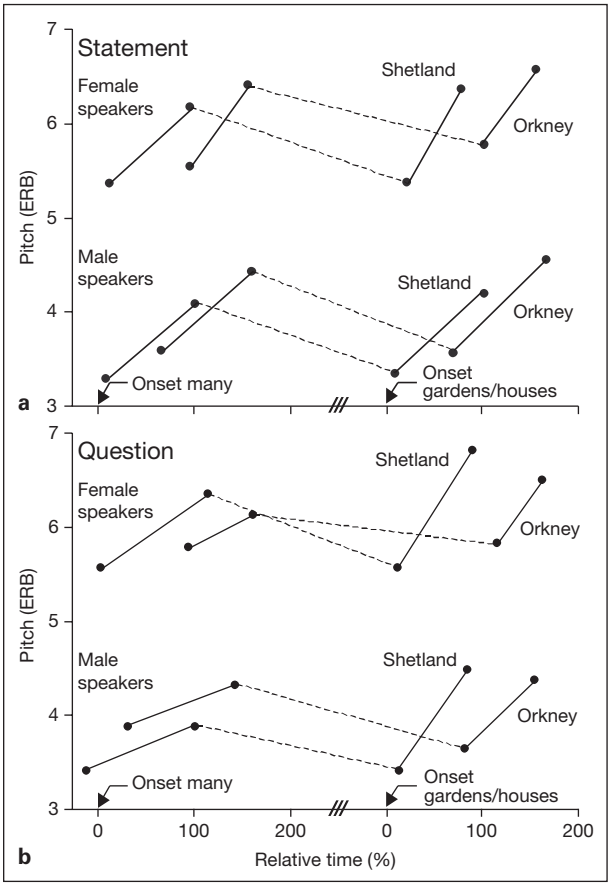

Fig. 3. Typical rise patterns for statements (**a**) and questions (**b**), conflated over two utterances, and broken down by speaker dialect and by gender. See text for further explanation.

gender or the dialect of the speaker. In the question versions of these sentences, the first accent (i.e. on *many*) is considerably smaller than its counterpart in the statement, while the second accent (i.e. on *gardens*/*houses*) is nearly twice as large, even when expressed in ERB, as the first. It has been shown for other languages (such as Dutch) that the sizes of successive accents increase over the course of an utterance in questions but not in statements [van Heuven and Haan, 2000; Haan, 2002]. In this respect, Orkney and Shetland intonation patterns do not differ from each other or from other languages.

One final parameter that seems to differentiate systematically between Orkney and Shetland intonation patterns is the alignment of the rise. It can be observed, very clearly, that the rise on *many* in Shetland dialect coincides with the first syllable, whereas both the onset and offset of the rise are located at a considerably later point in time for Orkney speakers, such that the peak is located at the syllable following the stress. The same shift in pitch-peak alignment is observed in the rise on *gardens*/*houses*.

The pitch and alignment data were analysed by separate one-way ANOVA for male and female speakers, assuming dialect (Orkney or Shetland) as a fixed factor. The results of these analyses are presented in table 4.

Table 4. Results of the one-way ANOVA (degrees of freedom, F ratio, η^2) for male and female speakers, with dialect (Orkney or Shetland) as a fixed factor

Acoustic parameter	Men			Women		
	df_1, df_2	F ratio	η^2	df_1, df_2	F ratio	η^2
R1 onset (alignment)	1,86	36.2**	0.294	1,48	326.9**	0.874
R1 peak (alignment)	1,86	139.1**	0.621	1,48	35.4**	0.429
R2 onset (alignment)	1,85	134.3**	0.611	1,59	383.1**	0.881
R2 peak (alignment)	1,85	196.5**	0.697	1,59	103.3**	0.628
R1 onset (ERB)	1,86	22.2**	0.205	1,48	2.58	0.052
R1 peak (ERB)	1,86	10.7*	0.112	1,48	<1	
R2 onset (ERB)	1,85	6.2*	0.074	1,59	6.8*	0.102
R2 peak (ERB)	1,85	<1		1,59	<1	

The η^2 statistic describes the proportion of total variance attributable to a factor. Mean data of 18 Orkney and 19 Shetland subjects. *$p \leq 0.05$; **$p \leq 0.001$.

When we examine table 4, we notice that there is a considerable difference between the values for male and female speakers. [It should be noted that the group sizes are not equal: 22 men (11 Orcadians and 11 Shetlanders) versus 15 women (7 from Orkney and 8 from Shetland). Consequently, only the values for eta squared (η^2) allow a straightforward comparison.] For men, the effect of dialect is strongest for peak alignment, while the F values are relatively low for rise-onset alignment. As for pitch level, the effect is largest for R1 onset, whereas the values for the other three pitch variables are very small and, at most, only significant at the 0.05 level. For women, the effect of dialect is markedly higher for rise-onset alignment than for peak alignment, nevertheless the effects for peak alignment are within the same range as for men. The effect of dialect appears to be almost negligible for pitch level. Looking again at figure 3 above, we see that the onset of the pitch rises of both R1 and R2 is located roughly 50 percentage points later for Orkney women than for Orkney men, while the peaks are located at about the same point in time, relative to the first-syllable duration. Consequently, the pitch excursions are somewhat steeper in female than in male speech. No such gender differences are observed for Shetland dialect. In conclusion, we may note that early versus late pitch rise alignment seems to be an important parameter, alongside low versus high mean pitch in the case of male speech, for differentiating between the speech melodies of Orkney and Shetland. Additionally, a multivariate analysis (results not presented here), with speaker dialect, sentence type (statement or question) and gender as fixed factors showed a significant effect of dialect (as in table 4), an effect of gender on rise onset alignment, interacting with dialect, and no effects or interactions of sentence type.

4.3.2 The Role of Mean Pitch and Pitch-Peak Alignment

In order to estimate how successful the two parameters (pitch-peak alignment and mean pitch) could be as perceptual cues to the dialect difference, Linear Discriminant Analysis (LDA) [Klecka, 1980] was applied. LDA finds an optimal linear combination of weighted parameter values that allows the separation of data points in pre-given categories. An LDA was set up to categorise the 148 utterance tokens into Orkney and Shetland dialect on the basis of the two parameters that were found to characterise the

Table 5. Weights associated with acoustic predictors in three LDAs, and percentage of correct classification for each discriminant function

Parameter	Discriminant function		
	pitch	alignment	all parameters
R1 onset (Z-ERB)	1.003		0.264
R1 peak (Z-ERB)	–0.266		–0.217
R2 onset (Z-ERB)	0.566		0.282
R2 peak (Z-ERB)	–0.552		–0.231
R1 onset (alignment)		0.111	0.140
R1 peak (alignment)		0.410	0.368
R2 onset (alignment)		0.592	0.524
R2 peak (alignment)		0.620	0.689
% correct Orkney	64.6	95.4	95.4
% correct Shetland	69.6	100	100
% correct all	67.2	97.8	97.8

difference between the two dialects. Since the pitch differs very strongly between male and female speakers, absolute pitch values (in ERB) were z-transformed within the set of male speakers (across the two dialects) and within the set of female speakers (also across the two dialects) separately.

We ran three separate LDAs. In the first analysis, the four F_0 parameters were used to categorise the dialect of the speaker. These are the z-normalised ERB values of onset and peak of the pitch rise on the first (R1) and second (R2) accents in each of the 148 utterances. In the second LDA, only the alignment parameters were included. These are the locations in relative time, expressed as a percentage of first syllable duration of the onset and peak of R1 and R2. In the final analysis, the pitch and alignment parameters were combined, yielding a set of eight predictors. Since only two categories (Orkney or Shetland) have to be discriminated, the LDA yields a single discriminant function, which is a sum of weighted normalised parameter values. Table 5 presents the standardised weights that are associated in each of the three analyses with each of the acoustic parameters that were entered.

It is apparent from the results of the LDA that correct classification of the two dialects is moderate when only the frequency values of the onsets and peaks of the rises R1 and R2 are used as predictors. Correct classification is 65% for Orkney and 70% for Shetland, i.e. no more than 20% above chance level (which is 50%). The strongest predictors within the set of four frequency values is provided by the onsets of the rises, especially R1, as is evidenced by the larger value of the weight coefficient for these parameters.

Dialect classification is almost perfect with the set of alignment parameters, with percentage correct scores between 95 and 100%. The strongest alignment predictors are associated with R2 alignment, both onset and offset. Examination of the incorrectly predicted Orkney utterances revealed that these concerned three instances of 'misaccented' pronunciations. In all three cases, the stress-accent was realised on *Bergen* rather than on *gardens* or *houses*, which resulted in an earlier alignment of the R2 onset, but with the pitch peak still delayed until well into the syllable following the stress.

Combining the pitch and alignment parameters does not increase percent correct identification; the same three Orkney tokens are incorrectly predicted. By far the best predictors of the set of eight are, once again, R2 onset and offset alignment.

4.4 Conclusions

In experiment 3, we found that, both in statements and in questions, the overall pitch level in Orkney was substantially higher than in Shetland. Moreover, it was established that there is a difference in pitch-peak alignment between the two dialects. This difference can be characterised as a shift in the location of the entire rise, i.e. both the onset and the peak. Shetland has early alignment, whereas the accent-lending rise in Orkney is late, such that in disyllabic words with initial stress the pitch peak does not occur on the stressed syllable, but is delayed to the unstressed syllable immediately following the stress. The outcome of an LDA of the measurement results indicated that the difference between Orkney and Shetland dialects is accounted for primarily by pitch-peak alignment, while high versus low overall pitch is only a weak predictor.

5 Experiment 4: Peak Alignment versus Overall Pitch Level

The outcome of the acoustic analysis presented in the previous section, however, does not in itself prove the perceptual relevance of any of the prosodic parameters. The chief aim of our final experiment, therefore, is to establish the role of pitch-peak alignment versus overall pitch level in the identification of the two dialects by native Orkney and Shetland listeners.[6]

5.1 Method

Using an experimental design similar to that of experiment 2, the utterance *There are many gardens in Bergen* was presented to native Orkney and Shetland listeners. The sentence was digitally recorded by 2 male speakers, both about 40 years old, one from Orkney (Kirkwall) and the other from Shetland (West Mainland); the recordings were the same as those used in experiment 2.

Two speech conditions were generated: (1) LPC-resynthesised speech (intelligible) and (2) buzzed speech (unintelligible; cf. experiment 2). In both speech conditions 16 different stimuli, i.e. 8 (intonation manipulations) $\times$ 2 (dialects) were generated. The following main types of test utterance were created: (1) early alignment (the pitch peak is located on the stressed syllable); (2) mid alignment (the pitch rise starts about halfway through the stressed vowel, such that the peak is located at the junction of the stressed and the following unstressed syllable); (3) late alignment (the pitch peak is shifted to the unstressed syllable following the stress).

It should be noted that the difference in timing between early, mid and late peak alignment is unequal. The early peak was invariably implemented exactly 120 ms after the onset of the stressed vowel, whereas both the mid and late peaks were located at specific landmarks in the segmental structure of the target word. For both speakers, the mid-aligned peak on *gardens* was implemented so that it coincided with the syllable junction between *gar-* and *-dens*; the late-aligned peak was realised just before the onset of the final fricative. As there was a difference in temporal organisation between the speakers, the mid and late peaks occurred at 90 and 250 ms later than the early peak for the Orkney

[6]Additionally, we investigated the effect of the number of pitch peaks per utterance and also the exact shape of the pitch rises, because whilst investigating Orkney and Shetland speech materials for our study reported in section 2, differences were observed with respect to these aspects of the intonation contour. Spontaneous Orkney speech seems to have distinct rise-fall pitch peaks on almost every word of any given utterance, while Shetland speech has pitch movements on accented words only. These complications in the experimental design have been omitted in the present article. The reader is referred to van Leyden [2004, 2006] for the complete report.

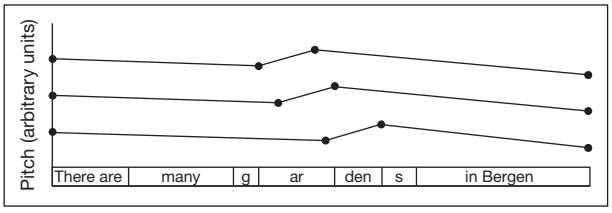

Fig. 4. Early, mid and late alignment. (Shetland segmental basis.)

speaker and at 100 and 210 ms later for the Shetland speaker. (Relative to Shetland speakers, Orcadians have short stressed initial syllables. For details see van Leyden [2004].)

Each of the manipulations as described here occurred in two versions: relatively low-pitched, starting at 100 Hz, as is normally found for this type of short utterances produced in isolation by Shetland speakers, and high-pitched, starting at 120 Hz, i.e. 'Orkney level'. Finally, for comparison purposes, 4 (2 dialects × 2 pitch levels) monotonised test utterances were also included. Monotonisation was done by changing the pitch contour into a flat line. The declination rate for all stimuli was set at four semitones per second, which is about the same as the original rate of the recorded utterances. The pitch manipulations in the stimuli were specified in terms of semitones for the sake of convenience; at the time the stimuli were generated, pitch target values in Praat could be entered either in Hertz or in semitones but not in equivalent rectangular bandwidth. Within the (male) pitch range of our manipulations the semitone and ERB scales are virtually equivalent.

Peaks consisted of a straight rise of five semitones relative to the declination line; the duration of the rise was set at 160 ms. The slope of the fall lasted until the end of the utterance. In all, 2 (pitch heights) × 4 (timings, including monotonised utterances) = 8 melodically different versions were generated on two segmentally different bases (Orkney, Shetland) yielding a set of 16 stimulus types. The pitch contours are illustrated in figure 4.

The manipulated speech fragments were converted from digital (16 kHz, 16 bit) to analog, allowing a signal bandwidth of just less than 8 kHz, and then recorded onto minidisc using a Sony MZ-R35 minidisc recorder. The stimuli were organised into two blocks, one speech condition per block, with fragments randomised within each block: (1) unintelligible and (2) intelligible. In order to compensate for a possible learning effect, two counterbalanced lists were created for each block. As in experiment 2, each test utterance was presented twice successively, with an interval of 1 s separating the two tokens. The interstimulus interval was 3.5 s, offset of second token to onset of first token of the next stimulus; an alert tone was played 1 s prior to each stimulus onset.

5.2 Subjects and Procedure

Forty listeners took part in the experiment, 20 from Orkney (12 male and 8 female) and 20 from Shetland (9 male and 11 female). The subjects were chosen from several parishes throughout Orkney and Shetland and were between 30 and 50 years of age. They reported no hearing problems and were not paid for their participation. None of the listeners had taken part in any of the other listening experiments reported above.

The block of intelligible speech (LPC-resynthesised) was always presented last in order to prevent listeners from transferring information gathered from this block to the unintelligible block (buzz). Each block was preceded by two practice fragments; responses to these trials were not included in the analysis. The experimental procedure was identical to the one described in section 2.2 above.

5.3 Results

A total of 40 (subjects) × 8 (intonation manipulations) × 2 (dialects, i.e. Orkney or Shetland segments) × 2 (speech conditions, i.e. buzz or LPC-resynthesised) = 1,280 responses were collected; there were no missing responses. The data were submitted to an RM-ANOVA with speech condition, speaker dialect (segments), peak timing and pitch height as within-subject factors and with dialect of listener as a between-subjects factor; degrees of freedom are Huynh-Feldt-corrected. Figure 5 presents the breakdown

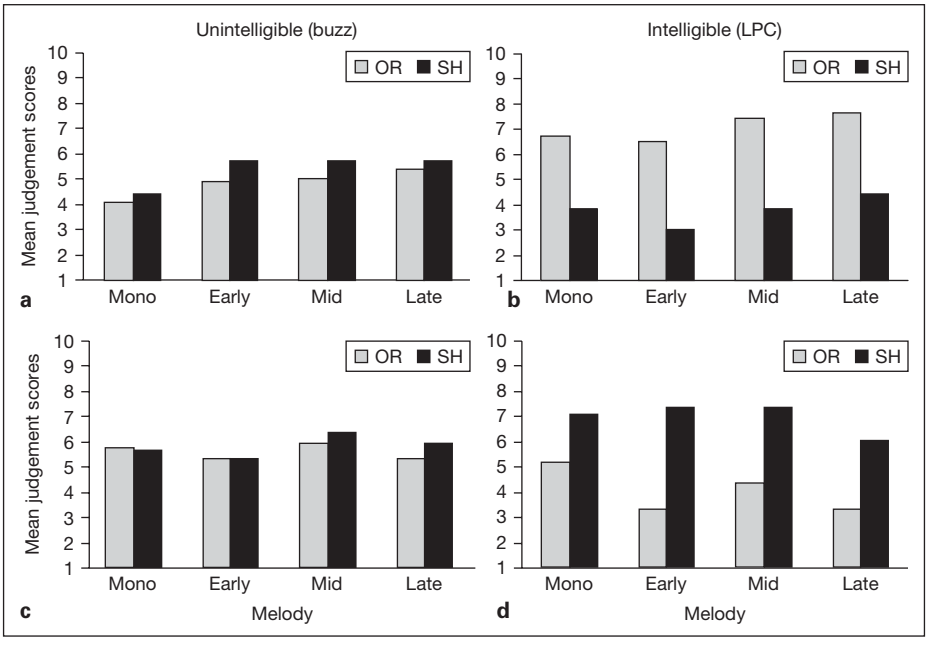

Fig. 5. Mean judgement scores (1 'definitely elsewhere'; 10 'definitely my island') for unintelligible speech (**a**, **c**) and intelligible speech (**b**, **d**); Orkney listeners top (**a**, **b**), Shetland listeners bottom (**c**, **d**). The scores are broken down by melody (monotonous, early, mid, late) and segmental information (OR, SH). Mean data of 20 Orkney and 20 Shetland subjects.

Table 6. Summary of RM-ANOVA for experiment 4

Effect/interaction	df_1	df_2	F	p
Speech condition (C)	1	38	0.031	=0.862
Segments (S)	1	38	0.211	=0.649
Melody (M)	3	79.2	2.7	=0.070
Listener group (L)	1	38	2.0	=0.162
Pitch height (P)	1	38	4.6	=0.038*
C × M	3	114	3.2	=0.028*
S × L	1	38	54.3	<0.001*
M × L	2.1	79.2	6.7	=0.002*
C × S × L	1	114	2.7	=0.051*

All effects and the significant interactions have been listed.

of mean judgement scores in four panels as in figure 2 above, but now with early, mid and late peak timing instead of Shetland (early) and Orkney (late) alignment along the x axis; the results are conflated over pitch height (low, high). A summary of the ANOVA is given in table 6.

We may note that out of five main effects only that of pitch height reaches significance. Generally, it would seem that the structure of the results reflects that of experiment 2 but all effects and interactions are substantially smaller. Large effects now show up as

Table 7. Summary of RM-ANOVA for four parts of data (fig. 5)

Effect/interaction	df_1	df_2	F	p
A. Orkney listeners, unintelligible				
Segments (S)	1	19	5.4	=0.032
Melody (M)	2.7	51.4	5.3	=0.004
B. Orkney listeners, intelligible				
Segments (S)	1	19	54.8	<0.001
Melody (M)	2.5	46.7	5.7	=0.004
C. Shetland listeners, unintelligible				
Pitch (P)	1	19	9.0	=0.007
D. Shetland listeners, intelligible				
Segments (S)	1	19	37.9	<0.001
Melody (M)	2.7	51.8	4.7	=0.007
Pitch (P)	1	19	4.7	=0.042
S × M	3.0	56.5	3.7	=0.017

moderate ones, and many moderate effects are now trends, at best. We will discuss the five significant effects and interactions separately.

The lower-pitched stimuli (5.8) were judged to be more typical of the 'own island' than the higher-pitch versions (5.3). The asymmetry is caused by the responses given by the Shetland listeners, who clearly associate lower pitch with Shetland, and high pitch with non-nativeness, whilst Orcadians have no such bias. The interaction between pitch level and listener group remains just short of significance, $F(1, 38) = 3.8$ ($p = 0.059$) and is therefore a trend at best.

Since all other main effects were insignificant, whereas the interactions between speech condition and melody, as well as between segments and listener group, melody and listener group, and the three-way interaction were all significant, we ran separate two-way RM-ANOVAs for each of the four panels in figure 5. The results of these analyses are presented in table 7.

In figure 5a, for Orkney listeners, monotonous versions are judged to be less native than the tonal versions, which do not differ from each other (Scheffé post hoc test, $\alpha < 0.05$). Shetland segments, even when made unintelligible, are judged to be a little more native overall than Orkney segments, even though the listeners are from Orkney (we will leave this effect undiscussed).

In figure 5c, for Shetland listeners, there appears to be a small advantage for Shetland segments (even when unintelligible) as well as for 'mid' alignment, with the pitch peak at the very end of the stressed syllable. However, these effects are statistically insignificant.

In the intelligible conditions there are – predictably – large interactions between the origin of the segments (Shetland, Orkney) and the listener group (table 7). Figure 5b shows that for Orkney listeners, regardless of the segmental base, the two later peak alignments are preferred over earlier alignment (but do not differ from each other by the Scheffé test). Early alignment is judged to be as poor as (or even poorer than) no melody at all.

Figure 5d shows that late peak alignment is disfavoured by Shetland listeners, when the segments are recognisably of Shetland origin. With Orkney segments, Shetland listeners prefer no melody over any tonal pattern. Within the tonal patterns, the mid alignment is judged to be the least non-native (Scheffé test) – which squares with the result of figure 5c.

5.4 Conclusions

The results obtained in experiment 4 indicate that there is a weak effect (in fact, no more than a statistical trend) for Shetland listeners to perceive low-pitched speech as being 'more Shetland' than high-pitched speech, whereas for Orkney listeners the overall pitch level is apparently not an important cue. In *unintelligible* speech, the effects of peak alignment as well as of segmental information are negligible for both listener groups. When presented with *intelligible* speech, both Orkney and Shetland listeners clearly (and predictably) differentiate between the two dialects based on segmental information. Alignment, too, is an important cue here, inasmuch that Orkney listeners regard early-aligned speech as 'less Orcadian' than late-aligned (the later the better), whereas Shetlanders consider late-aligned speech to be 'less Shetland' than early-aligned.

The fact that the effects in the present experiment are smaller than those of experiment 2, which basically targets the same stimulus dimensions, may be explained as follows. In experiment 4, the materials were controlled for alignment, pitch configuration as well as declination slope. In experiment 2, we used original pitch contours, which differed simultaneously with respect to all these features. Consequently, important prosodic cues that allow listeners to differentiate between the dialects could well have been eliminated. Moreover, in experiment 2 contours were either Orkney or Shetland, but never 'in between', as were many of the manipulations used in the present experiment. We feel that such a design made the situation in experiment 2 rather more transparent to the listener and thus resulted in a clear-cut response pattern.

6 Summary and Discussion

The aim of the present study was to find experimental support for impressionistic claims that there are intonational differences between Orkney and Shetland dialects. In experiment 1, we first explored the extent to which native listeners are able to distinguish between the two varieties on the basis of prosodic cues. The outcome of our perceptual study indicated that listeners distinguished quite clearly between the two intonational systems when they were presented with unintelligible speech samples in both dialects, i.e. when only melodic and temporal information was available. However, the two varieties were indistinguishable when listeners heard speech that was both monotonised and rendered unintelligible (i.e. LP-filtered). (Similarly, Cohen and 't Hart [1970] found that, when presented with unintelligible, monotonous speech, listeners were unable to distinguish between English and Dutch.) Furthermore, it was shown that the two dialects also differ distinctly with respect to their segmental structure. Our results are in line with other studies reporting that listeners are perfectly able to recognise language varieties solely on the basis of melodic information. We conclude that the prosodic difference between Orkney and Shetland speech is a matter of intonation rather than temporal organisation.

The purpose of experiment 2 was to establish the relative contribution of intonation and segmental information to the mutual identifiability of the dialects concerned. The crucial results of this experiment revealed that the contribution of segments and intonation to the identifiability of a speech sample was roughly equal. Nevertheless, for Shetland listeners, segmental deviations contributed more to non-nativeness than did a deviant intonation pattern – an effect that has commonly been reported for this type of study. However, Orkney listeners seemed to attach slightly more weight to intonation

than to segments. This finding contradicts previous work suggesting that melodic differences are always secondary cues in the identification of language varieties.

The aim of our acoustic investigation (experiment 3) was to isolate potential cues by which native Orkney and Shetland listeners differentiate between the two dialects. Two major differences between the prosodic systems of the dialects were observed. Firstly, it was found that, for both male and female speakers, the overall pitch in Orkney was substantially higher than in Shetland. Secondly, we found a significant difference with respect to the temporal alignment of the accent-lending rise-fall contours. In Shetland, an accent-lending pitch rise begins at the onset of the stressed syllable, while the pitch peak is located just before the offset of the same syllable. In Orkney speech, the temporal structure of the accent-lending rise differs, such that in disyllabic words with initial stress the start of the rise coincides with the offset of the stressed syllable and the H target is now delayed to the unstressed syllable immediately following the stress. An LDA clearly indicated that, acoustically, the difference between Orkney and Shetland dialects rests with the alignment parameter rather than with the pitch values.

In experiment 4, the relevance of the established prosodic parameters was evaluated perceptually by systematically manipulating the pitch pattern as well as the overall pitch level. The results of this investigation were found to correlate very well with our earlier perceptual data. Shetland listeners regard early peak alignment as characteristic of Shetland dialect whereas Orcadians attribute late alignment to Orkney. Overall pitch level played a minor role, but only for the Shetland listeners; it played no role at all for Orkney listeners. This outcome indicates that among the prosodic parameters alignment is indeed the stronger perceptual cue that allows the two listener groups to single out their fellow dialect speakers.

Although the overall difference in pitch level proved to be no more than a secondary cue in the distinction between Shetland and Orkney speech, we should point out nevertheless that this represents an unusual phenomenon. It has been shown in the literature that languages and language varieties may differ from each other in terms of mean pitch. For instance, the mean pitch of English and of Chinese was found to be higher than that of Dutch. The reason for the difference in these instances lies in the larger pitch range used in English (and even more so in Chinese). The lows of English intonation are lower than their counterparts in Dutch, but the highs are much higher than in Dutch [Willems, 1982; de Pijper, 1983; Gussenhoven, 2004; Chen, 2005]. Due to the asymmetry in the expansion of the pitch range in English (and Chinese) relative to Dutch the mean pitch in English assumes a higher value than that of Dutch. In the comparison of Shetland and Orkney (cf. fig. 3) we observe roughly the same pitch range for both varieties when plotted in equivalent rectangular bandwidth, however, both the highs and the lows in Orkney speech are approximately 0.3 ERB higher than in Shetland. We exclude the possibility that the higher mean pitch in Orkney is due to unrepresentative sampling. The distribution of mean pitch over the speakers in our samples contains no signs of outliers. Also, the difference in mean pitch was found to function as a cue in the perceptual contrast between Orkney and Shetland melody (experiment 4). It could never have functioned as a secondary cue if the difference had been a matter of accident. It is also important to note that the higher mean pitch in Orkney is found irrespective of speaker gender. Shifts in (mean) pitch have been reported in the literature that seem to be employed to emphasize differences in gender among the members of a linguistic community. It has often been noted that Japanese women speak at much higher pitches than women in other societies, and that Japanese

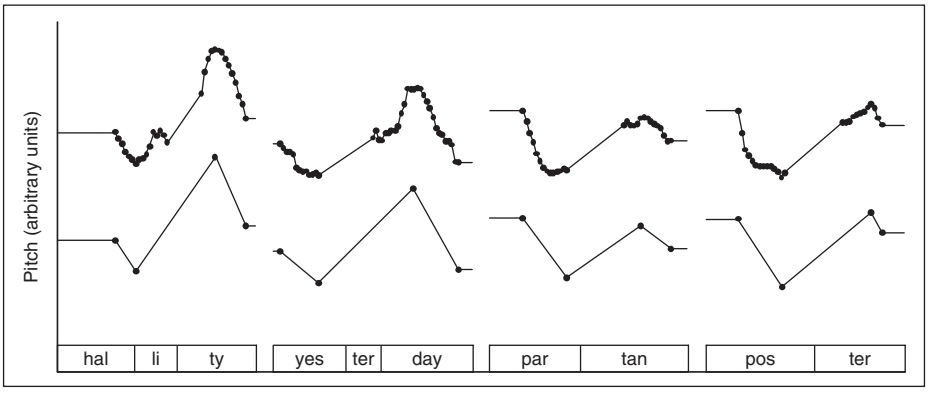

Fig. 6. Two- and three-syllable words with stress on the first syllable spoken by the same male speaker of Orkney dialect (hallity 'light-headed'; partan 'edible crab').

men have much lower pitches than their counterparts in other societies. Van Bezooijen [1993] noted that in the Dutch-speaking community, women from Flanders (Belgium) have considerably higher mean pitch than women in the western and northern parts of the Netherlands, even when the subjects were matched for body size. In this research, too, the difference could be ascribed to gender role.

The main conclusion of this paper is that the closely related dialects of Orkney and Shetland differ significantly with respect to their melodic structure. Apart from the secondary difference in mean pitch, Shetland speech has early pitch-peak alignment, whereas the pitch peak in Orkney is late, such that in disyllabic words with initial stress the pitch peak does not occur on the stressed syllable, but is delayed to the unstressed syllable immediately following the stress.

There are at least two possible views on the melodic difference in accent marking between Shetland and Orkney dialect. The difference could be interpreted as a matter of different phonetic implementations of the same underlying configuration of melodic targets H*L. In Shetland, the configuration has an early alignment with the H* locked to the stressed syllable; in Orkney, the entire configuration would then be time-shifted such that the H* coincides with the very end of the syllable following the stress. A consequence of the phonetic interpretation of the time-shift would be that not only the H* target but also the pre-accentual L would then be delayed by roughly a syllable. The alternative view would be phonological, whereby the accent-lending configuration would be analysed as L*H in Orkney speech, which would differ fundamentally from its H*L counterpart in Shetland. If indeed the Orkney accent has L*H, we should observe, first of all, that the low target remains locked to the stressed syllable, irrespective of the number of unstressed syllables that follow in the word. The post-accentual H part of the configuration could then very well be separable and move to a later position in longer words. Informal observation of three-syllable words with initial stress suggests that the L* targets remain stable but that the H target of the L*H pattern in Orkney dialect is delayed not just to the syllable following the stress but in fact to the very end of the word. This is illustrated in figure 6, which shows the raw and stylised F_0 contours on sentence-medial two- and three-syllable words with stress on the first

syllable pronounced by the same speaker from Orkney. (The four words occurred in four different sentences spoken during an informal interview unrelated to the present investigation.) It is quite clear from the figure that the pitch minimum is locked to the middle of the stressed syllable, but that the peak coincides with the end of the word. The Orkney pattern reminds us of the situation in Bornholm Danish. In this variant, the L* is locked to the stressed syllable while the H is time-locked to the last unstressed syllable in the stress group, and the pitch pattern is compressed or expanded in accordance with the composition of the stress group [Grønnum, 1990]. In the sentences, the target words occur at the end of a phonological phrase, so that possibly the H could also be interpreted as a boundary tone (which would not be part of the tonal specification of the word itself but of the next-higher domain, i.e. the phrase).

The above discussion is reminiscent of Dalton and Ní Chasaide's [2005] recent work on Donegal (late alignment) versus Connaught (early alignment) Irish; these authors, too, assume phonological status for the contrast rather than analysing the late alignment as a displaced H*.

In view of the above, we are inclined to interpret the melodic difference between Shetland and Orkney accentuation as a phonological difference between a high accent (H*) and a low accent (L*), respectively. However, our present study did not aim to be more than a contrastive phonetic investigation and therefore the materials we collected were not designed to test formally the predictions we derived from the phonetic versus phonological account of the melodic difference. In order to be able to decide on how the observed melodic differences in accent marking between the dialects of Orkney and Shetland should be interpreted phonologically, we need to investigate the systemic relations of peak synchronisation within the two dialects. This will be a concern for future work.

At this stage, little is known about the origin of the prosodic differences between Orkney and Shetland speech. Local lore has it that Orkney's lilting intonation is 'Norwegian' and, indeed, relatively late rises are also a feature of certain dialects spoken in Eastern Norway and Western Sweden. In these dialects too 'the main accent is followed by a rise extending to the next accent or to the end of the sentence. The impression is one of rising intonation for both statements and questions, a feature which is very striking to foreigners' [Gårding, 1998]. That all statements sound like questions is also generally mentioned by Shetlanders when they are asked to give some comments on Orkney speech. However, given the historical links between the Northern Isles and *western* Scandinavia, it is quite surprising that one should find *eastern* Scandinavian intonation in Orkney, but not in Shetland dialect, which appears to be the most 'Scandinavian' of the two varieties in terms of temporal organisation and lexis [van Leyden, 2002, 2004].

The linguistic situation as found in the Northern Isles is by no means unique in the world. Closely related dialects that differ prosodically have also been reported for language contact areas such as the Raja Ampat archipelago (Indonesia) [Remijsen, 2001] and Ireland. In fact, the state of affairs in Ireland is similar to the one found in the Northern Isles. As stated above, delayed peaks are also found in Donegal (Ulster Irish) and Belfast (Ulster English), while not only varieties of Irish spoken in Connaught (south of Donegal), but also Dublin English have early-aligned peaks [Grabe, 2002; Grabe and Post, 2002; Dalton and Ní Chasaide, 2003]. On account of this, Dalton and Ní Chasaide [2003] assume that late peak alignment, as found in Belfast and Glasgow dialects is, in the first case, a direct influence of Ulster Irish and in the second case, an

indirect influence. Note, however, that late rises have also been reported for languages that are completely remote from Gaelic or Scandinavian influence. In Southwest German varieties (Baden, Württemberg and Switzerland), for example, we find an intonation pattern that is very reminiscent of the Orkney pattern, whereas other varieties of German have a pattern more similar to the one found in Shetland and Eastern Scotland [Gilles, 2005].

To return to the Northern Isles, here we clearly have a language-contact situation too, with Norn and Lowland Scots spoken alongside each other for several centuries, and with the Gaelic-speaking Highlands nearby. Since, once Viking power waned in the region, Orkney became increasingly involved with mainland Scotland, one might speculate it was in fact Gaelic – rather than Scandinavian – influences that appear to have indirectly affected Orcadian speech melody (perhaps via the Caithness dialect). Concerning these matters, we agree with Barnes [1991] that a careful examination of the historical developments in the Northern Isles would be needed to gain more insight into the sociolinguistic circumstances surrounding the shift from Norn to Scots.

The findings of a diachronic study carried out by Hognestad [2006] seem to indicate that a shift from H* to L* prosody can occur within only a few generations. Similarly, a shift from H* to L* took place over a generation or two in Standard Copenhagen Danish very recently [Grønnum, 1992]. Likewise, the arrival in post-Viking Orkney of immigrants from mainland Scotland might well have initiated a similar prosodic change in the local dialect. Much research remains to be done, both on the melodic structure of Orkney and Shetland speech as well as on the possible similarities between the two dialects and the languages that might have affected them at some stage, i.e. Norwegian, Gaelic and Highland English.

Acknowledgement

We are very grateful to Klaus Kohler and to our reviewers, Nina Grønnum and Francis Nolan, for their useful comments. Thanks are also due to the (necessarily anonymous) Orkney and Shetland informants for their patience, encouragement and kindness.

Appendix: The Four Intonation Contours Used in Experiment 2

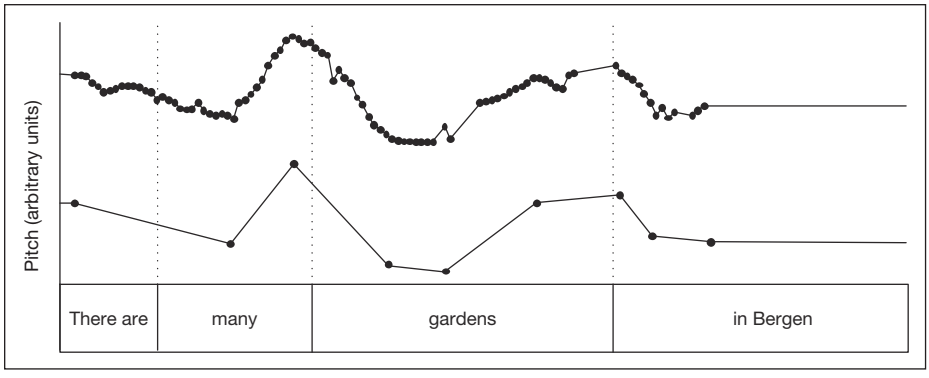

Fig. 7. Original and stylised intonation contour of speaker or1 (with creaky voice in 'Bergen', hence no pitch points).

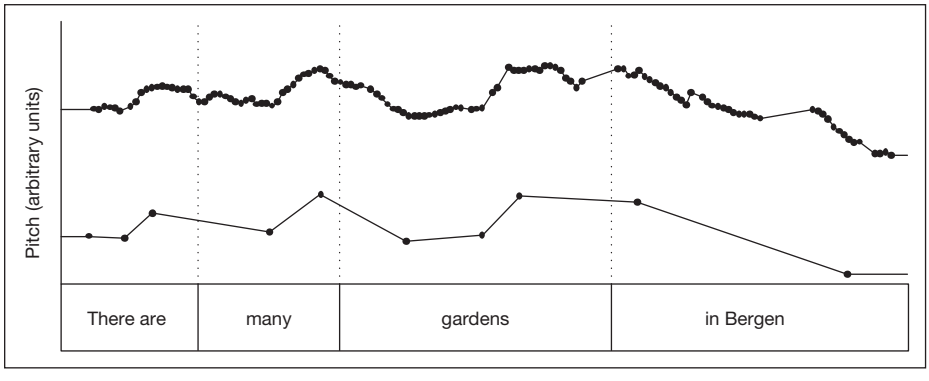

Fig. 8. Original and stylised intonation contour of speaker or2.

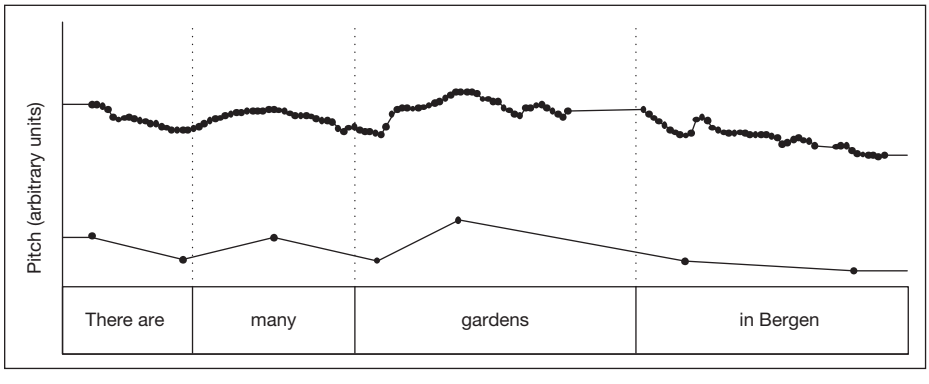

Fig. 9. Original and stylised intonation contour of speaker sh1.

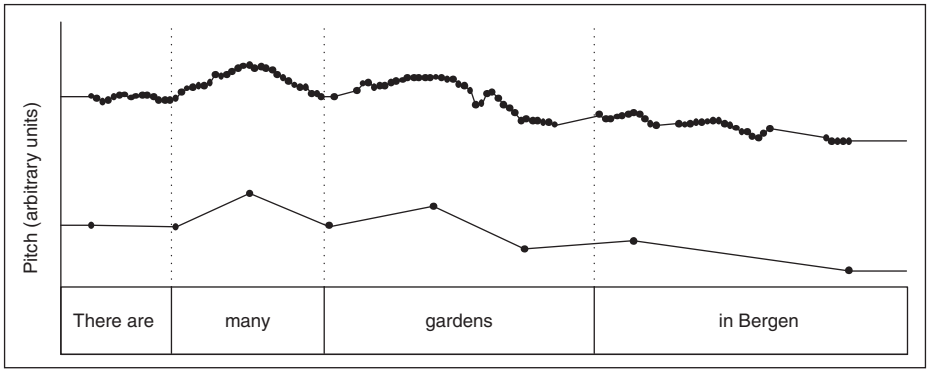

Fig. 10. Original and stylised intonation contour of speaker sh2.

References

Barnes, M.P.: Reflections on the structure and the demise of Orkney and Shetland Norn; in Sture Ureland, Broderick, Language contact in the British Isles (Linguistische Arbeiten 238), pp. 429–460 (Niemeyer, Tübingen 1991).

Bezooijen, R. van: Verschillen in toonhoogte: natuur of cultuur? (Differences in pitch: nature or culture?) Gramma/TTT *2:* 165–179 (1993).

Biemans, M.: The effect of biological gender (sex) and social gender (gender identity) on the three pitch measures; in van Bezooijen, Kager, Linguistics in the Netherlands 1998, pp. 41–52 (Benjamins, Amsterdam 1998).

Boersma, P.; Weenink, D.: Praat: doing phonetics by computer (version 4.3) (http://www.praat.org, 2005).

Catford, J.C.: Shetland dialect. Shetland Folkbook *3:* 71–76 (1957).

Chen, A.: Universal and language-specific perception of paralinguistic intonational meaning; doct. diss. Radboud Universiteit, Nijmegen, LOT Dissertation Series 102 (LOT, Utrecht 2005).

Cohen, A.; 't Hart, J.: Comparison of Dutch and English intonation contours in spoken news bulletins. IPO Annu. Progr. Rep. *5:* 78–82 (1970).

Cruttenden, A.: Intonation; 2nd ed. (Cambridge University Press, Cambridge 1997).

Dalton, M.; Ní Chasaide, A.: Modelling intonation in three Irish dialects; in Solé, Recasens, Romero, Proc. 15th Int. Congr. Phonet. Sci., pp. 1073–1076 (Casual Productions, Barcelona 2003).

Dalton, M.; Ní Chasaide, A.: Tonal alignment in Irish dialects. Lang. Speech *48:* 441–464 (2005).

Gårding, E.: Intonation in Swedish; in Hirst, Di Cristo, Intonation systems, pp. 112–130 (Cambridge University Press, Cambridge 1998).

Gilles, P.: Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung. Linguistik – Impulse und Tendenzen 6 (de Gruyter, Berlin 2005).

Gooskens, C.: On the role of prosodic and verbal information in the perception of Dutch and English language varieties; doct. diss. Catholic University Nijmegen (1997).

Gooskens, C.; Bezooijen, R. van: The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties; in Berns, van Marle, Present-day dialectology, problems and findings, pp. 173–192 (Mouton de Gruyter, Berlin 2002).

Grabe, E.: Variation adds to prosodic typology; in Bel, Marlin, Proc. Speech Prosody 2002 Conf., pp. 127–132 (Laboratoire Parole et Langage, Aix-en-Provence 2002).

Grabe, E.; Post, B.: Intonational variation in English; in Bel, Marlin, Proc. Speech Prosody 2002 Conf., pp. 343–346 (Laboratoire Parole et Langage, Aix-en-Provence 2002).

Grønnum, N.: Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German. Phonetica *47:* 182–214 (1990).

Grønnum, N.: The groundworks of Danish intonation: an introduction (Museum Tusculanum Press, Copenhagen 1992).

Gussenhoven, C.: The phonology of tone and intonation (Cambridge University Press, Cambridge 2004).

Haan, J.: Speaking of questions: an exploration of Dutch question intonation; doct. diss. Catholic University Nijmegen, LOT Dissertation Series 52 (LOT, Utrecht 2002).

Hart, J. 't; Collier, R.; Cohen, A.: A perceptual study of intonation: an experimental-phonetic approach to speech melody (Cambridge University Press, Cambridge 1990).

Henton, C.: Fact and fiction in the description of female and male pitch. Lang. Commun. *9:* 299–311 (1998).

Hermes, D.; Gestel, J.C. van: The frequency scale of speech intonation. J. acoust. Soc. Am. *90:* 97–102 (1991).

Heuven, V.J. van; Haan, J.: Phonetic correlates of statement versus question intonation in Dutch; in Botinis, Intonation: analysis, modeling and technology, pp. 119–144 (Kluwer, Dordrecht 2000).
Heuven, V.J. van; Leyden, K. van: A contrastive acoustical investigation of Orkney and Shetland intonation; in Solé, Recasens, Romero, Proc. 15th Int. Congr. Phonet. Sci., pp. 805–808 (Casual Productions, Barcelona 2003).
Hognestad, J.K.: Tonal accents in Stavanger: from western towards eastern Norwegian prosody? In Bruce, Horne, Nordic prosody IX, pp. 107–116 (Lang, Frankfurt am Main 2006).
Johnston, P.: Regional variation; in Jones, The Edinburgh history of the Scots language, pp. 433–513 (Edinburgh University Press, Edinburgh 1997).
Klecka, W.R.: Discriminant analysis (Sage Publications, Beverly Hills 1980).
Leyden, K. van: The relationship between vowel and consonant duration in Orkney and Shetland dialects. Phonetica *59:* 1–19 (2002).
Leyden, K. van: Prosodic characteristics of Orkney and Shetland dialect: an experimental approach; doct. diss. Leiden University, LOT Dissertation Series 92 (LOT, Utrecht 2004).
Leyden, K. van: Pitch-peak alignment vs. overall pitch level in the identification of Orkney and Shetland dialects; in Bruce, Horne, Nordic prosody IX, pp. 175–184 (Lang, Frankfurt am Main 2006).
Leyden, K. van; Heuven, V.J. van: Prosody versus segments in the identification of Orkney and Shetland dialects; in Solé, Recasens, Romero, Proc. 15th Int. Congr. Phonet. Sci., pp. 1197–1200 (Casual Productions, Barcelona 2003).
Markel, J.D.; Gray, A.H.: Linear prediction of speech (Springer, Berlin 1976).
Moulines, E.; Verhelst, W.: Time-domain and frequency-domain techniques for prosodic modification of speech; in Kleijn, Paliwal, Speech coding and synthesis, pp. 519–555 (Elsevier Science, Amsterdam 1995).
Nooteboom, S.G.: The prosody of speech: melody and rhythm; in Hardcastle, Laver, The handbook of phonetic sciences, pp. 640–673 (Blackwell, Oxford 1997).
Peters, J.; Gilles, P.; Auer, P.; Selting, M.: Identifying regional varieties by pitch information: a comparison of two approaches; in Solé, Recasens, Romero, Proc. 15th Int. Congr. Phonet. Sci., pp. 1065–1068 (Casual Productions, Barcelona 2003).
Pijper, J.R. de: Modelling British English intonation, an analysis by resynthesis of British English intonation; doct. diss. Utrecht University (1983).
Remijsen, B.: Word-prosodic systems of the Raja Ampat languages; doct. diss. Leiden University, LOT Dissertation Series 49 (LOT, Utrecht 2001).
Schaeffler, F.; Summers, R.: Recognizing German dialects by prosodic features alone. Proc. 14th Int. Congr. Phonet. Sci., San Francisco 1999, pp. 2311–231.
Thorsen, N.: An acoustical investigation of Danish intonation. J. Phonet. *6:* 151–175 (1978).
Willems, N.: English intonation from a Dutch point of view (Foris, Dordrecht 1982).