A Contrastive Acoustical Investigation of Orkney and Shetland Intonation

Vincent J. van Heuven and Klaske van Leyden
Phonetics Laboratory, Universiteit Leiden Centre for Linguistics (ULCL)
E-mail: V.J.J.P.van.Heuven@let.leidenuniv.nl, K.van.Leyden@let.leidenuniv.nl

ABSTRACT

This study presents an acoustical investigation of the melodic differences between Orkney and Shetland dialects. Typical melodic profiles of short statements and (yes/no) questions were derived for Orkney and Shetland. Differences were found in overall pitch (Orkney > Shetland, both valley and peak pivot points), and in the temporal alignment of the accent-lending rise-fall contours, which are located on the stressed syllable in Shetland but clearly shift to the next, i.e. post-stress, syllable in Orkney. Linear Discriminant Analysis was used to determine which of the two sets of cues affords the best separation of Orkney versus Shetland intonation patterns. Pitch level as a single parameter yielded no more than 70% correct. Peak alignment allowed 96% correct dialect classification; a mere 1%-increment was obtained when pitch level was added to the alignment cue. The results of this acoustic study correlate almost perfectly with our earlier perceptual data.

1. INTRODUCTION

The variety of Lowland Scots spoken on the islands of Orkney and Shetland – north of mainland Scotland – is a relatively recent development based on a shared parent language, Norn, which was introduced to the islands by Viking settlers from Norway. However, present-day Orkney and Shetland dialects differ markedly in their prosodic make-up. Differences in temporal structure between the two dialects were studied by [1]. In [2] it was demonstrated that native listeners proved quite able to distinguish Orkney from Shetland speech on the basis of the intonation contour only. We now present an acoustical investigation of the melodic differences between the Orkney and Shetland varieties. A systematic comparison of the acoustical measurements with the perceptual results in [2] may allow us to isolate potential cues in the melodic systems by which listeners differentiate the two dialects.

2. METHOD

The materials consisted of a corpus of 148 utterances, comprising tokens of two declarative sentences and two yes-no questions: *There are many gardens*/*hooses in Bergen* and *Are there many gardens*/*hooses in Bergen?*[1] The sentences were read out by eleven male and seven female speakers from Orkney and eleven male and eight female speakers from Shetland (i.e. 37 speakers all together); all speakers were aged between 30 and 50 years of age. The stimuli were printed individually on cue cards and presented one at a time; the presentation order differed per speaker. Subjects were instructed to read at a natural speaking rate and to pronounce the sentences with a pitch accent on *gardens*/*hooses*. Each speaker produced the entire set of sentences twice. This resulted in 296 recorded sentences: 37 (speakers) x 4 (stimuli) x 2 (repetitions). However, for each of the subjects, only the second recording of the material was acoustically analysed. Only when a speaker mispronounced a particular sentence, which happened in about five tokens, was the first recoding used.

Using PRAAT [3] speech-processing software, F_0 was extracted (autocorrelation method) for each utterance. Rise-fall configurations of the pitch accent on *gardens*/*hooses* 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 [4] using PSOLA analysis and resynthesis [5]). 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 of the vocalic nuclei within the syllables) with which the pitch movements were aligned. No values were entered in those cases where the expected rise-fall accent was omitted from a particular constituent by the speaker.

3. RESULTS AND DISCUSSION

Figure 1 presents the basic intonation patterns that were realised on the accents on *many* and *gardens*/*hooses*, with pitch plotted in ERB as a function of linear time, in separate panels for statements (top panel) and questions (bottom panel). The time scale is in seconds such that the 0 point has been chosen to coincide with the onset of the first accented vowel, i.e. the onset of /e/ in *many*. In each panel, the intonation patterns of the male speakers are seen in the bottom half of the pitch range, typically between 3 and 5

[1] As in other varieties of Lowland Scots, *house* is pronounced as /hus/ in both Orkney and Shetland, hence, the dialect spelling *hoose*.

ISBN 1-876346-48-5

ERB, whilst the female contours are located in the upper half of the range, typically between 5 and 7 ERB. Only the rise portions of the accents have been plotted. The dotted line connecting the end of the rise on *many* to the onset of the rise on *gardens*/*hooses* ignores the – highly variable – location of the endpoint of the pitch fall somewhere between the two rises.

The first difference between Orkney and Shetland patterns that strikes us in figure 1 is that, both in statements and in questions, the overall pitch in Orkney is substantially higher than in Shetland. Given the rather 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 appears that within the sexes the speakers of each dialect community are rather tightly clustered. 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 due to infelicitous speaker sampling. Rather, we would surmise that the overall higher pitch in Orcadian is a feature of the dialect.

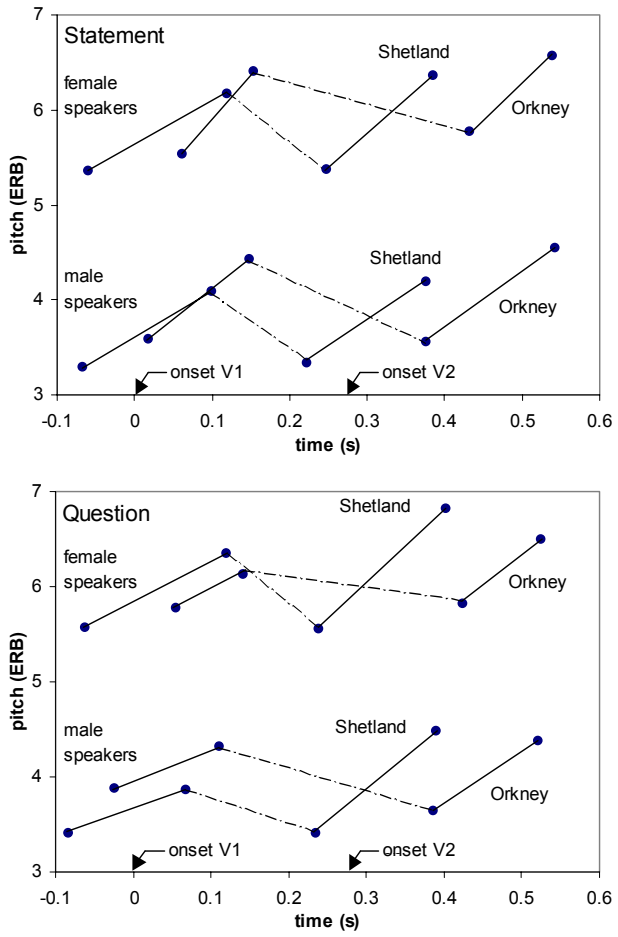

Figure 1. Typical rise patterns for statement (top panel) and question (bottom panel) broken down by dialect and by sex. The dotted lines indicate the assumed slope of the fall. See text for further explanation.

Secondly, we observe that, both in Orkney and in Shetland, the intonation contours have the same shapes (and excursion sizes, at least when expressed in ERB) for men as they have for women. So, it seems that the only difference that is related to sex of speaker is the overall (i.e. mean) pitch.

Third, in the statements the pitch rises are roughly equally large (excursion sizes between 0.8 and 0.9 ERB) for *many* and for *gardens*/*hooses*, irrespective of the sex or the dialect of the speaker. In the question versions of these sentences, the first accent (on *many*) is considerably smaller than its counterpart in the statement, while the second accent (in *gardens*/*hooses*) is roughly twice as large (even when expressed in ERB) as the first. It has been shown for other languages (notably Dutch) that the sizes of successive accents increase over the course of the sentence in questions but not in statements [4, 6]. In this respect Orkney and Shetland intonation do not differ from each other or from other languages.

There is one final parameter that seems to differentiate systematically between Orkney and Shetland intonation patterns. It can be observed, very clearly, that the rise on *many* in Shetland dialect straddles the CV-boundary between /m/ and /e/ of *many*, whereas onset of the rise is located after the CV boundary in Orkney. The relatively early alignment (pitch rise begins at the syllable onset) is what we normally find for languages such as English or Dutch [7, 8]. Orkney has the accent-lending rise, and presumably the entire rise-fall contour, shifted to a later position, such that the pitch peak does not occur on the stressed syllable but is delayed to the unstressed syllable following the stress. This alignment shift has also been found for a number of Danish dialects, including Copenhagen Danish [9]. The same shift in segmental alignment is observed in the rise on *gardens*/*hooses*. Therefore, the conclusion seems warranted that early versus late alignment is an important parameter, next to low versus high mean pitch, which differentiates between the speech melodies of Shetland versus Orkney.

In order to estimate how successful the two parameters, mean pitch and segmental alignment, could be as perceptual cues to the dialect difference, Linear Discriminant Analysis [10, 11] was applied. LDA finds an optimal linear combination of weighted parameter values that allows the separation of data points in pre-given categories. In our problem, the LDA was set up to categorise the 148 sentence tokens into Shetland and Orkney dialect on the basis of just the intonation parameters that were found to characterise the difference in figure 1, i.e., the pitch values of the onset and end-point (i.e. peak) of the rise on *many* (R1) and on *gardens*/*hooses* (R2), and the time coordinates of the pitch points expressed relatively to the vowel onset of the stressed syllable that the rise is associated with. 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 separately. For the relative timing of the pitch rises, R1 and R2, a different normalisation procedure was adopted. The onset of the stressed vowel was given the value of 0% relative time, whilst the end of the stressed rhyme (i.e., of /u/ in *hooses* and /ɑr/ in *gardens*) was set at

ISBN 1-876346-48-5

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% reflects a rise the onset of which occurs at a point in time that finds itself half the length of the stressed rhyme before the vowel onset. Similarly, a rise offset at 200% relative time is located two rhyme lengths after the onset of the stressed vowel.

We ran three separate LDAs. In the first analysis, only 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 sentence token. In the second LDA, only the alignment parameters were included, i.e. the location in relative time (expressed in percent rhyme duration) of the onset and peak of R1 and R2. In the third analysis, the pitch and alignment parameters were combined, yielding a set of eight predictors.

Since only two categories have to be discriminated, the LDA yields a single discriminant function, which is a sum of weighted normalised parameter values. Table 1 presents the weights that are associated in each of the three analyses with each of the acoustic parameters that were entered.

Table 1. Weights associated with acoustic predictors in three Linear Discriminant Analyses, and percent correct classification for each discriminant function.

	Discriminant function		
Parameter	**Pitch only**	**Align only**	**Both**
R1 onset (Z-ERB)	1.003		0.335
R1 peak (Z-ERB)	– 0.266		– 0.206
R2 onset (Z-ERB)	0.566		0.225
R2 peak (Z-ERB)	– 0.552		– 0.258
R1 onset (alignment)		0.159	0.160
R1 peak (alignment)		0.565	0.490
R2 onset (alignment)		0.460	0.442
R2 peak (alignment)		0.627	0.685
% correct Orkney	66.7	94.4	94.4
% correct Shetland	69.7	97.4	98.7
% correct all	68.2	95.9	96.6

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. Percent correct stops at 67 for Orkney and at 70 for Shetland, i.e., it does not amount to more than 20 percent above chance level (which is 50%). The strongest predictors within the set of four frequency values is provided by the onsets the rises, especially R1, as is evidenced by the larger value of the weight coefficient for these parameters.

Much better classification is afforded by the set of alignment parameters, with percent correct scores between 94 and 97. In fact, if the LDA is carried out for the statement utterances only, correct classification is 100% (results not presented here). The strongest alignment predictors are associated with the pitch peaks of R1 and R2.

Combining the pitch and alignment parameters allows just a marginal improvement over the alignment-only analysis. Percent correct is not increased for the classification of Orkney intonation, and a mere 1 percent better classification is obtained for Shetland.

4. CONCLUSIONS

In the present study we measured characteristic pitch/time coordinate values for the pitch patterns realised on the accented words in two lexically different sentences produced by 18 speakers of Orkney and 19 speakers of Shetland dialect. The results indicated that basically the same melodic patterns were used by the male and female speakers of each dialect. The only sex-related difference was one of overall pitch. This allowed us to normalise for the effect of sex and to come up with a unified account of the characteristic difference between the sentence melodies of the two dialects.

Two sets of parameters were isolated, viz. (i) the pitch values at the onset and at the peak of the accent-lending rises, and (ii) the corresponding segmental alignment of these pitch points. Although both types of parameter seem to differentiate quite well between the two dialects when we visually compare the (averaged) profiles (see figure 1), a token-individual statistical analysis (LDA) unequivocally indicates that the difference between the two dialects resides with the alignment parameters rather than with the pitch values. In fact, the alignment parameters, especially those associated with the accent peaks, afford near-perfect assignment of the utterances to the two dialects, whilst dialect classification did not exceed 70% correct (a mere 20% above chance) when only the pitch parameters were used as predictors. Adding the pitch parameters to the alignment parameters resulted in no more than 1 percent improvement of the dialect classification.

The alignment difference between the two language varieties can be characterised as a shift in the location of the entire rise, i.e. both the onset and the peak. The Shetland alignment is “early” as is normal in English and Dutch. The accent-lending rise in Orkney is “late” and typically shifts so as to align with the unstressed syllable following the accent, a pattern that has also been found in a number of Scandinavian language varieties.

The present study allows a straightforward account of the results of our perception study [2], in which we found that Orkney and Shetland listeners distinguished quite well between the speech melodies of their native dialect and that of other varieties. Figure 2 plots the mean judgement score (1 ‘definitely from elsewhere’ to 10 ‘definitely own dialect’) plotted as a function of the strongest acoustic parameter differentiating between Orkney and Shetland intonation, i.e. the relative timing of the accent peak. Of course, the mean judgement scores in figure 2 were computed only for those stimuli in which the original segmental information was replaced by a buzz-like sound (sawtooth wave) that perfectly mimicked the timing and amplitude envelope of the original speech, while obliterating the spectral information. Therefore, the dialect authenticity judgements were exclusively based on the prosodic cues in the stimuli.

ISBN 1-876346-48-5

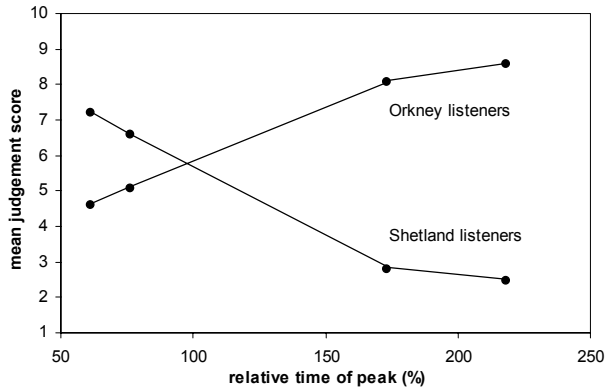

Figure 2. Mean judgement score (1: 'definitely from elsewhere' – 10: 'definitely own dialect') plotted as a function of relative timing of the accent peak, broken down by listener group.

The data are shown for the Orkney and Shetland listener groups separately. The leftmost two dots are the mean peak alignment values for the two Shetland speakers whose utterances were the basis of the stimuli for the perception experiment. The rightmost two dots represent the peak alignment values for the two Orkney speakers. We observe, first of all, the very large effect of the geographic origin of the speakers, showing that early peak alignment is deemed characteristic of Shetland dialect and late alignment is strongly associated with Orkney. But, crucially, also within the two pairs of speakers, the individuals are ordered such that the peak alignment corresponds with the authenticity of the intonation pattern. So, the effect of peak alignment is systematically found both between and within the speaker groups. Mean peak alignment correlates with the dialect judgment at $r = -0.978$ for the Shetland listeners, and at $r = 0.989$ for the Orkney listeners ($N = 4$, $p < 0.05$). This, to us, convincingly indicates that the alignment is the strongest perceptual cue by far that allows the two listener groups to single out their fellow dialect speakers – even though the ultimate proof can only be given by an experiment systematically manipulating the accent alignment along a continuum running between early to late, synthesized on a single segmental utterance.

Finally, one might be tempted to conclude that Orkney dialect is influenced by its Scandinavian substrate (Norn). Yet, such a conclusion might be false. If the melody of Norn should have survived until the present day, it is difficult to understand why the typically Scandinavian temporal organisation got lost in Orkney, while it is still found in Shetland [1], which does not boast the "Scandinavian" melody. The delayed rises may not be unique to Scandinavian languages and dialects; given that not only Orcadians, but also Welsh (and other Gaelic) speakers are commonly described as having lilting (sing-song) intonation, the phenomenon of delayed peaks might be responsible for this auditory impression. If so, delayed peaks might be common to both Scandinavian and Gaelic.

Acknowledgment

We gratefully acknowledge the financial support provided for the second author by the *Leids Universiteits Fonds* (LUF).

REFERENCES

[1] K. van Leyden, "The relationship between vowel and consonant duration in Orkney and Shetland dialects." *Phonetica* 59, pp. 1-19, 2002.

[2] K. van Leyden and V.J. van Heuven, "Prosody versus segments in the identification of Orkney and Shetland dialects." This volume.

[3] P. Boersma and D. Weenink, "PRAAT: a system for doing phonetics by computer," *Report of the Institute for Phonetic Sciences Amsterdam*, 132, 1996.

[4] V.J. van Heuven and J. Haan, "Phonetic Correlates of Statement versus Question Intonation in Dutch," in *Intonation: Analysis, Modelling and Technology*, A. Botinis, Ed., pp. 119-144. Dordrecht/Boston/London: Kluwer, 2000.

[5] E. Moulines and W. Verhelst, "Time-domain and frequency-domain techniques for prosodic modification of speech," in: *Speech Coding and Synthesis*, W.B. Kleijn and K.K. Paliwal, Eds., pp. 519-555. Amsterdam: Elsevier Science, 1995.

[6] V.J. van Heuven and J. Haan, "Temporal development of interrogativity cues in Dutch," in: C. Gussenhoven and N. Warner, Eds., *Papers in Laboratory Phonology VII*, pp. 61-86. Mouton de Gruyter: Berlin, 2002.

[7] J. Caspers and V. J van Heuven, "Effects of time pressure on the phonetic realisation of the Dutch accent lending pitch rise and fall," *Phonetica* 50, pp, 161-171, 1993.

[8] J. Caspers, *Pitch movements under time pressure. Effects of speech rate on the melodic marking of accents and boundaries in Dutch*, Holland Institute of Generative Linguistics dissertation series No. 10. The Hague: Holland Academic Graphics, 1994.

[9] N. Thorsen, N. "A study on the intonation of sentence intonation – evidence from Danish," *Journal of the Acoustical Society of America*, 67(3), pp. 1014-1030, 1980.

[10] W.R. Klecka, *Discriminant Analysis*. Beverly Hills/ London: Sage Publications, 1980.

[11] V.J. van Heuven, J. Haan and J.J.A. Pacilly, "Automatic recognition of sentence type from prosody in Dutch," *Proceedings of Eurospeech '97*, pp. 1483-1486. Rhodes, 1997.