Prosody versus segments in the identification of Orkney and Shetland dialects

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

ABSTRACT

This study examines the relative importance of intonation versus segmental structure to the (mutual) identifiability of Orkney and Shetland dialects by native listeners of each of these varieties. In both normal (intelligible) and filtered (unintelligible) speech, three types of test utterance were generated: original intonation, monotonised speech, and transplanted melodies. In experiment 1, listeners proved quite able to distinguish Orkney versus Shetland speech on the basis of intonation contour only. In experiment 2, it was shown that intonation contour was a more important cue than segmental information for Orcadians but not for Shetland listeners. This result defies earlier findings, and common-sense reasoning, suggesting that prosodic differences are always secondary cues in the identification of language varieties.

1. INTRODUCTION

The dialects now spoken on the islands of Orkney and Shetland – north of mainland Scotland – have evolved from a shared parent language, Norn, which was introduced by Viking settlers in the 9th century. This Scandinavian language remained the chief medium of communication in the Northern Isles for several centuries until it was finally abandoned in favour of Lowland Scots in the course of the 18th century.

Shetland, the northernmost of the two island groups, appears to have maintained its Norn substrate to a greater extent than Orkney. Dialect differences are apparent in the syllable structure [1] [2], the lexicon and morphosyntax. However, the most striking difference between Orkney and Shetland speech is the dissimilarity in intonation (i.e. speech melody). Impressionistically, Shetland dialect has a rather narrow pitch range, while Orkney speech has a relatively wide pitch range and is characterised by a very distinctive, lilting, 'rise-fall' intonation pattern.

Research – on English as well as on other languages – consistently shows that speech melody plays a *less* important role than segmental information in the identification of language varieties [3]. Challenging this generalisation, we tested the relative importance of intonation versus segmental information to the identification of Orkney and Shetland dialects (with Standard Scottish English, SSE, as a control condition) by native listeners of each of these varieties.

Two experiments will be reported. In the first, fragments of spontaneous speech were presented both with and without the original intonation contour (monotonised), in order to examine to what extent native Orkney and Shetland listeners are able to distinguish between the two dialects. In the second experiment, we examined the relative importance of intonation versus 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 realisations 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, which given the intonational pattern of an Orkney speaker, is identified as produced by an Orcadian, 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.

2. EXPERIMENT 1

2.1 Method

The main aim of this experiment is to examine the role of intonation in the identification of Orkney and Shetland dialects. Using a design similar to [3], short fragments of spontaneous speech were presented to native listeners of the two dialects. The speech fragments were about 12 seconds in duration and of semantically neutral content. The selected speakers were males; three speakers were from Orkney and three from Shetland; only one fragment per speaker was included. Two speech conditions were created: normal (intelligible) and low-pass filtered (unintelligible) speech. The LP-filtering was done by computer using PRAAT speech processing software [4]. The speech signal was low-pass filtered at 300 Hz, with a band smoothing of 50 Hz. In both normal and filtered speech two intonation conditions were generated: original speech melody and monotonised speech. Monotonisation was done with PRAAT software by changing the pitch contour into a flat line, using PSOLA analysis and resynthesis [5]. This line was given no declination, since the (distinguishing) role of declination in Orkney and Shetland dialects is not known. The speech fragments were monotonised at 100 Hz, which was approximately the mean pitch across all speakers.

The manipulated speech fragments were organised into four blocks: 1) LP-filtered & monotonised; 2) LP-filtered & intonation contour; 3) normal speech &

ISBN 1-876346-48-5 © 2003 UAB

monotonised; 4) normal speech & intonation contour. The stimuli were presented to twenty listeners, 10 from Orkney (6 male and 4 female) and 10 from Shetland (6 male and 4 female). The subjects were issued response sheets on which they were asked to indicate, for each utterance, where they thought that a particular speaker was 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 equilibrium, the listeners were forced to make a choice, however tentative.

2.2 Results

Figure 1 presents the mean judgement score broken down by speech condition and by dialect.

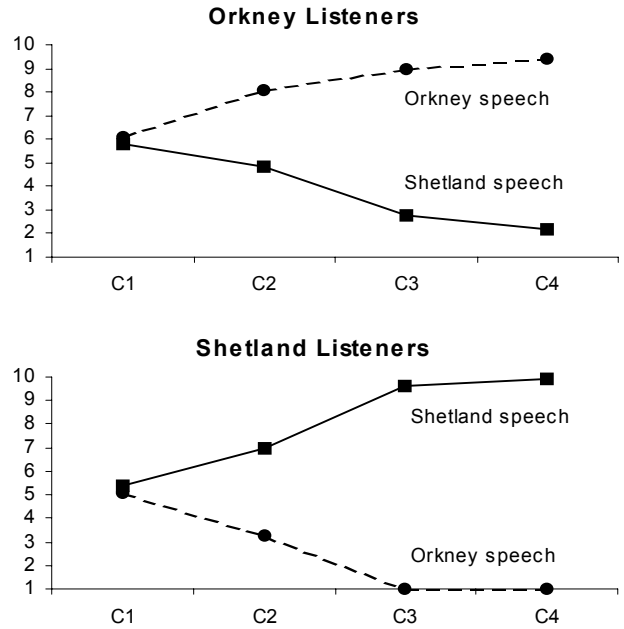

Figure 1. Mean judgement score broken down by speech condition and by dialect, for Orkney (top) and Shetland listeners (bottom). C1= LP-filtered & monotonous; C2= LP-filtered & original intonation contour; C2= normal speech & monotonous; C4= normal speech & original intonation contour.

As can be seen in figure 1, for Orkney listeners, there is a large effect of the dialect of the speaker, $F(1,231) = 313.8$ ($p < 0.001$), which strongly interacts with presentation condition, $F(3,231) = 42.8$ ($p < 0.001$), with judgement scores differentiating incrementally as a function of the linguistic cues that were 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 x condition interaction.

Clearly, then, condition C1 (monotonous and 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 (C2), the origin of the speakers is distinguished rather well. However, when monotonised intelligible speech is presented (C3) the differentiation between the varieties is better. Combining both information sources (C4) yields (nearly) maximal differentiation.

In this experiment the elimination technique proved suitable for demonstrating the role of intonation in distinguishing between dialects. However, it does not allow us to quantify the relative importance of intonation versus segmental information in the perceptual identification of the language varieties. In experiment 2, therefore, we examined the relative contribution of intonation versus 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.

3. EXPERIMENT 2

3.1 Method

This experiment examines the relative importance of intonation versus segmental information to the (mutual) identifiability of Orkney and Shetland dialects (with SSE as a control condition). The utterance *There are many gardens in Bergen* was presented to native Orkney and Shetland listeners. The sentence was recorded by two speakers from Orkney, two from Shetland and one SSE speaker from Central Scotland; all speakers were male. Three speech conditions were generated: 1) LPC-resynthesized speech (intelligible), 2) LP-filtered speech (unintelligible) and 3) buzz (unintelligible).

Using PRAAT software [4], a buzz was created by replacing segmental information by a spectrally invariant buzz-like sound (sawtooth wave) while preserving the amplitude and fundamental frequency modulations of the original, that is, all spectral information was removed from the speech signal but prosodic variation was maintained. This way, we can be certain that the perceptual identification of a particular language variety is based on prosodic cues only.

LPC-resynthesis was the done in order to conceal the identity of the speakers (in the small island communities chances were substantial that our listeners personally knew one or more of the speakers). Speech was analysed and resynthesized with five formants and associated band-widths in the 0 to 5 kHz range, with a sawtooth wave as the source signal. The resulting quality was highly intelligible but – at least in the short utterances used in our experiment – effectively masked the speaker's identity.[1]

For each of the speech conditions (buzz, LP-filtered, LPC-resynthesized), three types of test utterance were generated: original intonation, monotonous speech and transplanted melodies. The manipulations were done interactively using PSOLA analysis and resynthesis as implemented in PRAAT. Six transplantations were implemented: Orkney contours grafted onto the Shetland and the SSE utterances, Shetland contours imposed on the

[1] Listeners were debriefed after their participation in the experiment. None of the listeners could ever identify any of the speakers they had heard on the tape.

Orkney and SSE utterances and SSE contour on the Orkney and the Shetland utterances. For each original utterance F0 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 duration 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 manipulated sentences were organised into six blocks as follows: 1) buzz & monotonised; 2) buzz & (manipulated) intonation contour; 3) LP-filtered & monotonised; 4) LP-filtered & (manipulated) intonation contour; 5) LPC-resynthesized & monotonised; 6) LPC-resynthesized & (manipulated) intonation contour. The stimuli were randomised within each block and presented to 39 listeners, 19 from Orkney (10 male and 9 female) and 20 from Shetland (10 male and 10 female). The task of the subjects was the same as for experiment 1.

3.2 Results

The results of this experiment are presented in figure 2. The judgement scores (1: 'definitely from elsewhere', 10: 'definitely my island') are broken down by speech condition (buzz, filtered, resynthesized) and broken down further by intonation manipulation. These are the original but stylised contours for Orkney and Shetland (ORKork and SHsh) and for SSE (SSEsse). Manipulations also include two hybrid versions: Orkney segments with Shetland pitch contours (ORKsh) and Shetland segments with Orkney contours (SHork).[2, 3]

For Orkney listeners, there is a large effect of pitch manipulation. Overall, there is a clear split between the topmost two conditions and the bottom three. The former two conditions share the property that they have Orkney pitch contours. The bottom three end up with much lower scores since these have non-Orkney pitch contours. The main effect of pitch condition is significant, $F(4,953) = 192.0$ ($p < 0.001$). A post-hoc analysis for contrasts (Scheffé procedure) indicates that each condition differs significantly from all others ($p < 0.05$). There is a relatively small overall effect for speech condition (buzz, filtered, and resynthesized speech), $F(2,953) = 6.8$ ($p = 0.001$). Only the resynthesized and the buzzed speech differ in the post hoc analysis. However, there is considerable interaction between pitch manipulation and speech type, $F(8,953) = 9.5$ ($p < 0.001$) indicating that there is greater differentiation among the pitch manipulation in the resynthesis (with clearly intelligible segments) than in the two other conditions, where segmental information is largely obliterated.

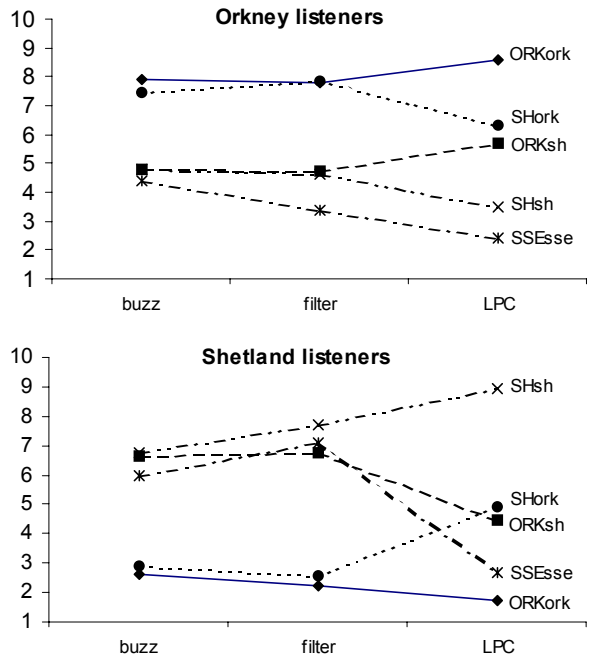

Figure 2. Mean judgement score broken down by speech condition and by dialect (manipulation), for Orkney listeners (top) and Shetland listeners (bottom). See text for explanation of labels.

The pattern of results for Shetland listeners is almost a mirror image of that for Orkney. Speech with Orkney pitch contours is discarded as non-Shetland, whereas, at least for the unintelligible speech conditions, SSE and Shetland pitch contours receive similar scores. Yet, when the segmental information becomes available, the scores for SSE drop sharply to within the non-Shetland range. The main effect of pitch condition is significant, $F(4,1005) = 266.8$ ($p < 0.001$). A post-hoc analysis for contrasts (Scheffé procedure) indicates that each condition differs significantly from all others ($p < 0.05$). The effect for speech condition (buzz, filtered, and resynthesized speech) is also significant, $F(2,1005) = 7.9$ ($p = 0.001$) but the speech types do not differ from each other in the post hoc analysis. Again, and crucially, there is considerable interaction between pitch manipulation and speech type, $F(8,1005) = 28.5$ ($p < 0.001$).

The judgement scores for the intelligible speech condition (resynthesized speech) are presented in more detail in table 1.

The scores indicate that Shetland listeners clearly differentiate between segmentally Shetland speech 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, the pattern is more complicated, as judgement scores are spread out across the entire range. Also, speech with Orkney pitch contours, but with Shetland segmental structure, is classified as Orcadian (i.e. ≥ 6.0).

When in original Orcadian 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

[2] In the designation of the pitch conditions 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.

[3] Due to space limitations, all manipulations from and to SSE had to be omitted from figure 2 (and from subsequent analyses and discussion).

ISBN 1-876346-48-5 © 2003 UAB

segments are replaced by their Shetland counterparts, the scores drop by 2.3 points. Clearly, then, the detrimental effect of replacing the intonation contour is larger (by 0.6 point) than that of replacing the segments.

Table 1. Judgement scores for resynthesized speech, broken down by pitch condition.

Pitch manipulation		**Origin listeners**	
segments	pitch	Orkney	Shetland
ORK	ork	8.6	1.8
ORK	mono	7.7	3.0
ORK	sh	5.7	4.4
SH	sh	3.5	8.9
SH	mono	4.9	8.0
SH	ork	6.3	4.9
SSE	sse	2.4	2.7

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).

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 influences, whether prosodic or segmental, than Orcadians do to Shetland speech.

4. DISCUSSION AND CONCLUSION

The first aim of this paper was to find experimental support for impressionistic claims that there are intonational differences between the dialects of Orkney and Shetland. Our first experiment shows that native listeners of Orkney and Shetland dialect distinguish quite clearly between the two intonational systems when they are presented with un-intelligible 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 monotonised as well as unintelligible (i.e. low-pass filtered) speech. Therefore, the conclusion is warranted that the prosodic difference is a matter of intonation rather than temporal organisation.

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

Our final, and most ambitious aim, was to present a case where closely related language varieties differ more strongly in their prosody than in their segmental make-up. We feel intuitively – but so far lack systematic data to support our idea – that dialects, or closely related language varieties, should always differ more in their segmental properties than in terms of intonation. For one thing, segments are produced at a rate of, say, between 10 and 20 per second, whilst even a simple pitch rise or fall (or combination of both) spans an entire syllable. Except in cases of abnormal tone crowding, then, the number of perceptually relevant pitch changes will be much less than 10 per second. Also, segmental information may differ along a larger variety of dimensions than pitch movements. Our results show that, at least for Orkney speech as judged by native Orcadians, foreign – i.e. Shetland – intonation detracts more from its acceptability than segmental information. Apparently, although the cards are stacked in favour of segmental cues, we have found one case of two closely related language varieties that differ more in their intonation than in their segmental properties.

At this stage, it is not yet clear how the divergence between Orkney and Shetland intonation has come about. The geographical situation (the sea separating the island groups) as well as the economical situation (an agrarian economy versus a sea-faring nation) undoubtedly played a role in the origin of the linguistic differences. As Orkney is very close to the Scottish mainland, it is also likely that Gaelic influences have indirectly (i.e. through Highland English with a Gaelic substrate) influenced the speech melody of Orcadian.

Acknowledgment

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

REFERENCES

[1] J.C. Catford, “Shetland dialect,” *Shetland Folkbook* 3, pp. 71-76, 1957.

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

[3] C. Gooskens and R. van Bezooijen, “The role of prosodic and verbal aspects of speech in the perceived divergence of Dutch and English language varieties,” in *Present-day Dialectology, Problems and Findings*, J. Berns and J. van Marle, Eds., Berlin: Mouton de Gruyter, 2002.

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

[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.