Dialect recognition in a noisy environment: preliminary data

Erik J. Eriksson and Kirk P. H. Sullivan

Department of Philosophy and Linguistics, Umeå University

Abstract

The ability to identify the dialect spoken by an individual by voice alone has been widely investigated for many languages. The studies suggest a large degree of individual variation in dialect recognition ability and that factors such as perceptual distance and competence in the language interact with this variation. Trained listeners have also been shown to be unstable in their judgements of dialect. In the study presented here naïve listeners attending a trade fair in Umeå were asked to select between eight given dialects when identifying two dialects. The test environment was noisey. The level of the background noise changed with the flow of people and activities associated with the trade fair. The noisy environment did not appear to affect the listeners responses.

Introduction

Many studies have examined the listener's ability to place a person's dialect and national background based on voice alone (e.g. Bayard et al., 2001; Bayard and Sullivan, 2005; Clopper and Pisoni, 2004; Clopper et al., 2006; Cunningham-Andersson, 1996; Doeleman, 1998; Markham, 1999; Preston, 1993; Remez et al., 2004; Sullivan and Karst, 1996; Williams et al., 1999). Together these studies shown that some listeners are able to place a speaker's regional origin with a reasonable degree of accuracy, yet that listeners use large dialectal areas that group together similar dialects. This grouping increases with perceptual difference from the listener's own dialect. These studies also point to a large degree of variation in listener ability.

Focusing on research on listeners' ability to discriminate, classify and judge the naturalness for recordings of Swedish dialects, variation in listener ability is also found. Cunningham-Andersson (1996) found variation between listeners in their ability to discriminate between and classify Swedish dialects. Yet, they also found that native – non-native listener background impacted upon their listeners' responses. Markham (1999), in an investigation of how well trained-listeners were able to identify a range of Swedish dialects, natural and imitated, found that some of these listeners were surprisingly unstable in their judgements of speaker background:

> . . . it became clear when collating the data that some judges were far less able to successfully identity dialectal accents or dialectal colouring in accents than the rest of the group. This was despite assurances from all judges that they had a good ear for Swedish accents. (p. 297)

When earwitnesses are asked to describe the voice they have heard, dialect forms a key part of the description (Hollien, 2002). Research by Schlichting and Sullivan (1997), and Zetterholm et al. (2002) have shown that voice imitation can lead to high levels of misidentification. The level of achieved accuracy has been shown not to be affected by a corrected hearing problem (Eriksson, Czigler, Skagerstrand and Sullivan, forthcoming). Further, working with a bidialectal speaker Eriksson, Schaeffler, Sjöström, Sullivan and Zetterholm (forthcoming) have shown the perceptual dominance of dialect when all other variables are controlled for by using a native bidialectal speaker. Their listeners based their speaker identification based on dialect.

This paper presents preliminary data from a pilot study that considered how naïve native Swedish listeners perform in a dialect identification task with uncontrolled background noise levels. The study included recording of Swedish dialects and imitations of Swedish dialects.

Table 1: The dialects selected by the listeners for each of the presented clips. S1 is presentation of the sentence "Ärten kom på museum", S2 is the presentation of the sentence "En riktig prinsessa", ST = Stockholm dialect, SC = Scanian, Bidialect = bidialectal speaker, imitation = the imitated dialect recordings, ST1 and ST2 = Stockholm speakers, and SC1 and SC2 = Scanian speakers.

	Bidialect ST		Bidialect SC		Imitation ST		Imitation SC		ST1		SC1		ST2		SC2	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Dalarna	1	0	1	1	4	0	0	0	2	1	1	1	2	0	0	0
Göteborg	1	3	0	2	0	0	2	2	1	0	2	1	0	2	0	4
Närke	4	0	0	2	3	3	2	0	2	0	4	3	7	3	1	3
Norrland	1	1	4	0	1	1	1	2	5	3	0	0	0	2	3	2
Skåne	3	0	4	4	1	1	8	6	1	1	5	5	1	1	8	3
Småland	1	4	3	5	2	1	5	5	1	1	3	7	1	2	3	3
Stockholm	4	7	2	2	4	7	0	1	3	8	0	0	2	4	2	1
Uppland	3	3	1	2	3	5	0	3	3	4	3	0	5	4	1	2

Method

Participants

The listeners

The participants were collected at a popular public trade in Umeå in Norther Sweden. People who approached Umeå University's Display were asked if they wished to take part in a perception experiment. Their participation was voluntary and without reimbursement. Twenty-two people completed the task. Four reported hearing a known hearing problem (Male: N=2, mean age= 42.5, age range=23-62; Female: N=2, ND=2) and 18 reported no hearing problem (Male: N=10, mean age = 34, age range = 7-66; Female: N=8, mean age = 28.63, age range = 17-39).

The speakers

The speakers were: two male speakers, SC1 (27 years of age) and SC2 (45 years old) who spoke the Scanian dialect of Swedish; two male speakers ST1 (25 years of age) and ST2 (28 years old) who spoke a variety of Stockholm Swedish; one male (36 years old) who spoke both the Scanian dialect and a variety of the Stockholm dialect (see Eriksson, Schaeffler, Sjöström, Sullivan and Zetterholm, forthcoming, for more information about this speaker) and one professional imitator (circa 45 years of age) who imitated both the Scanian and Stockholm dialects (see Zetterholm, 2003, for more information about this imitator).

Material

The two speakers of Scanian Swedish and two speakers of Stockolm Swedish read the fairy tail "The Princess and the pea". The bidialectal speaker read this fairy tail once in his Scanican accent and once in his Stockholm accent. From each of these recordings the following two sentences were clipped: "Ärten kom på museum" (The pea ended up in the museum) and "En riktig prinsessa" (a real princess). The professional imitator imitated the bidialectal speaker's readings of the two sentences. Thus, in total there were eight recordings of each sentence.

Procedure

The participants sat wearing head-phones in front of a laptop PC. The participants were first asked their place of birth, their current place of residence, their age, their gender and whether they had any hearing problem. Then, one by one, the sentences extracted from the recordings were played over headphones and the listeners asked to select the dialectal background of the speaker heard. Listeners made their choice through mouse clicks of one of the presented radio buttons. These buttons were labelled Dalarna, Göteborg, Närke, Norrland, Skåne, Småland, Stockholm and Uppland.

The dialects were randomized across the radio buttons for each played recording and upon presentation no default radio button was selected. This reduces impact of listeners getting bored and clicking the same button over and over again, or hitting enter (which results in a select-action on a radio button). Once the listeners had clicked a radio button the next sample was played and a new set of radio buttons drawn on screen. This was repeated for each of the 16 trials (8 recordings x 2 sentences). After the test was completed the results for the individual listener was presented

Table 2: The dialects selected by the listeners for each of the presented clips grouped into the regions Central, South and Other. S1 is presentation of the sentence "Ärten kom på museum", S2 is the presentation of the sentence "En riktig prinsessa", ST = Stockholm dialect, SC = Scanian, Bidialect = bidialectal speaker, imitation = the imitated dialect recordings, ST1 and ST2 = Stockholm speakers, and SC1 and SC2 = Scanian speakers.

	Bidialect ST		Bidialect SC		Imitation ST		Imitation SC		ST1		SC1		ST2		SC2	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
Central	12	10	4	7	14	15	2	4	10	13	8	5	16	11	4	6
South	4	4	7	9	3	2	13	10	2	2	8	12	2	3	11	6
Other	2	4	4	2	1	1	3	4	6	3	2	1	0	4	3	6

Table 3: The dialects selected by the listeners for each of the presented clips grouped for speaker and region. ST = Stockholm dialect, SC = Scanian, Bidialect = bidialectal speaker, imitation = the imitated dialect recordings, ST1 and ST2 = Stockholm speakers, and SC1 and SC2 = Scanian speakers.

	Bidialect ST	Bidialect SC	Imitation ST	Imitation SC	ST1	SC1	ST2	SC2
Central	22	11	29	6	23	13	27	10
South	8	16	5	23	4	20	5	17
Other	6	6	2	7	9	3	4	9

summarily. The test was run on a PC in a Java application custom designed for this test.

Results

The listeners who reported a hearing problem are not analysed due to their low number. The listener responses for the listners reporting no hearing known problem are presented in Table 1. The table shows the selected dialect(s) for each presented sentence. The responses for those with reporting no hearing problem and those reporting a hearing problem are presented separately.

Table 2 compresses the data presented in Table 2 into three areas. A Central Swedish group of dialects consiting of the dialects spoken in the regions of the regions of Dalarna, Närke, Stockholm and Uppland. A southern group of dialects consiting of the dialects spoken in the regions of Skåne and Småland. The remaining two dialects, one from the North of Sweden (Norrland) and one from the West of Sweden (Göteborg) are placed in the group Other. The Other group contains dialect areas that are distinct from those of spoken by the speakers.

As visual inspection of the responses reveals no major differences between the two sentences, it is, therefore, possible to merge the responses. Table 3 presents the sentence compressed responses.

Discussion

Visual inspection of Tables 1, 2 and 3 show a spread of listener responses for all of the presented voices with none of the voices being assigned the correct dialect or dialect region 100% of the time. From Tables 2 and 3, it can be seen that some of the clips were more easily placed than others. ST1 and ST2 are most frequently placed in the Central Group, whereas SC1 and SC2 have mixed placements.

Table 3 suggests that SC2 is less clearly placed than SC2, but upon examination of Table 2 it can been seen that SC1's "Ärten kom på museum" was poorly placed and SC2's "En riktig prinsessa" was poorly placed and that the difference suggested by Table 3 is an artefact of the summation rather than a reflection of a difference between how the listeners placed these clips. Of note is that it is not the phonetic information available in the sentence, per se, that results in lower accuracy rates as it is not the same sentence that resulted in the lower correct dialect region identification. The difference could be due to background noise variation when these stimuli were presented or due to fewer clear dialect clues being contained in SC1's "Ärten kom på museum" and SC2's "En riktig prinsessa".

All the dialect imitations were well placed in the broad dialect regions. The Stockholm imitations were marginally better placed than the Scanian imitations. Both dialect imitations were bet-

ter identified than the clips from the Stockholm and Scanian speakers (ST1, SC1, ST2 and SC2). This might be due to lower background noise when there clips were presented or due to the imitator having selected and focused on traits associated with these two dialects when he recorded these imitations.

The placing of the bidialect speakers four clips reveals a difference between his Stockholm and Scanian dialects. His Stockholm accent is more accurately placed than his Scanian accent. Again this could be due to differences in background noise, or to differences between the dialect information contained in the sentences. An acoustic and an auditory analysis of the bidialectal speakers clips undertaken by Eriksson et al. (forthcoming) found that these clips contained similarities even if they were perceived as two distinct dialects. It is possible that these similarities contribute to the lower correct dialect placement rate for the Scanian dialect or not.

There is a lower correct dialect placement rate for the Scania clips than the Stockholm clips (See the summary in Table 3) for all the voices. This could be due to the experiment being conducted in Northern Sweden. An detailed comparative analysis of the clips may permit clarification of which dialect cues could have been used by the listeners in this task.

The noisey environment does not appear to have adversely affected the listener' responses. They mimic those found in, for example, (e.g. Clopper and Pisoni, 2004; Clopper et al., 2006; Preston, 1993). This lack of impact is surprising. This pilot study needs to be repeated in controlled noisey environments that include a range of different realistic background noise types. The results from this pilot study suggest that some listeners can place a speaker's dialect into very large dialect areas, yet that others are unable to do this. Further research is needed to explain why this is so.

Acknowledgement

This research was funded by the grant "Imitated voices: A research project with applications for security and the law" Dnr K2002-1121:1-4 awarded to Umeå University by The Bank of Sweden Tercentenary Fund.

We thank Anders Mårtensson and Elisabeth Zetterholm for providing the imitated material.

References

Bayard D and Sullivan K P H (2005). A taste of kiwi: Does the Swedish palate differ in sensitivity and preference from the New Zealanders'? *Te Reo*, 48:29 – 40.

Bayard D, Weatherall A, Gallois C and Pittam J (2001). Pax Americana? accent attidutional evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*, 5:22 – 49.

Clopper C G, Levi S V and Pisoni D B (2006). Perceptual similarity of regional dialects of american english. *Journal of the Acoustical Society of America*, 119(1):566 – 574.

Clopper C G and Pisoni D B (2004). Effects of talker variability on perceptual learning of dialects. *Language and Speech*, 47(3):207 – 239.

Cunningham-Andersson U (1996). Learning to interpret sociodialectial cues. *Dept Speech, Music and Hearing - Quarterly Progress and Status Report, TMH-QPSR*, 37(2):155 – 158.

Doeleman R (1998). *Native reations to nonnative speech*, vol. 13 of *Studies in Multilingalism*. Tilburg, Netherlands: Tilburg University Press.

Eriksson E J, Czigler P E, Skagerstrand Å and Sullivan K P H (forthcoming). Effects of age and age-related hearing loss on speaker recognition or can senior citizens be reliable earwitnesses?

Eriksson E J, Schaeffler F, Sjöström M, Sullivan K P H and Zetterholm E (forthcoming). On the perceptual dominance of dialect.

Hollien H (2002). *Forensic Voice Identification*. San Diego, CA: Academic Press.

Markham D (1999). Listeners and disguised voices: the imitation and perception of dialectal accent. *Forensic Linguistics*, 6(2):289 – 299.

Preston D (1993). Folk dialectology. In D Preston, ed., *American Dialect Research*. Amsterdam – Philadelipha: John Benjamins Publishing Company.

Remez R E, Wissig S C, Ferro D F, Liberman K and Landau C (2004). A search for listener differences in the perception of talker identity. *Journal of the Acoustical Society of America*, 116(4, Pt. 2):2544.

Schlichting F and Sullivan K P H (1997). The imitated voice – a problem for voice line-ups? *Forensic Linguistics*, 4(1):148 – 165.

Sullivan K P H and Karst Y N (1996). Perception of English accented by native British English speakers and Swedish learners of English. In P McCormack and A Russell, eds., *Proceedings of the Sixth Australian International Conference on Speech Science and Technology*, 509 – 514. Adelaide, Australia: Australian Speech Science and Technology Association.

Williams A, Garrett P and Coupland N (1999). Dialect recognition. In D Preston, ed., *Handbook of Perceptual Dialectology, Volume 1*. Amsterdam – Philadelphia: John Benjamins Publishing Company.

Zetterholm E (2003). *Voice Imitation: A Phonetic Study of Perceptual Illusions and Acoustic Success*. Ph.D. thesis, Lund University, Lund, Sweden.

Zetterholm E, Sullivan K P H and van Doorn J (2002). The impact of semantic expectationon the acceptance of a voice imitation. In C Bow, ed., *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, 379 – 384. Melbourne, Australia: Astralian Speech Science and Technology Association.