# DIALECT MAPS AND DIALECT RESEARCH; USEFUL TOOLS FOR AUTOMATIC SPEECH RECOGNITION?

*Arne Kjell Foldvik* and *Knut Kvale***

*Dept. of Linguistics, NTNU, Trondheim, Norway. **Telenor R&D, Kjeller, Norway,

## ABSTRACT

Traditional dialect maps are based on data from carefully selected informants which usually results in clear-cut dialect borders, isoglosses, with one dialect characteristic present on one side of the isogloss and absent on the other.

We illustrate some of the problems and pitfalls connected with using dialect maps for ASR by comparing results from traditional dialect research with investigations of the Norwegian part of the European SpeechDat database, centred on the two main types of /r/ pronunciation.

Our analysis shows that traditional dialect maps and surveys may be of limited use in ASR. To what extent the Norwegian findings have parallels in other countries will depend on two main factors, dialect allegiance vs. a national standard pronunciation and the extent to which the population is sedentary or mobile. Results from traditional dialect research may therefore be more useful in ASR of other languages than Norwegian.

## 1. INTRODUCTION

Traditional dialect maps are based on data from carefully selected informants, ideally people who have lived in one area throughout their life. This selection of informants usually results in clear-cut dialect borders, isoglosses, with one dialect characteristic present on one side of the isogloss and absent on the other [1].

How useful are dialect maps for automatic speech recognition (ASR) purposes? It is attractive but simplistic to suppose that locating a speaker or a caller on the telephone would be sufficient for the speech recogniser to activate the appropriate acoustic models for the relevant dialect area. But it is unfortunately the case that traditional dialect maps give an idealised picture of the linguistic landscape, one where isoglosses delimit uniform linguistic communities, where there is also little or no variation in other aspects of pronunciation between speakers.

In this paper we illustrate some of the problems and pitfalls connected with using dialect maps by comparing results from traditional dialect research with investigations of /r/ pronunciations in a recently compiled database of Norwegian.

## 2. TYPES OF /r/ PRONUNCIATION

There are two main types of /r/ pronunciation in Norwegian, an apical tap, [ \ ], and a dorsal approximant or fricative, [ʋ]. As can be seen in Figure 1 and Figure 2 the apical and dorsal /r/ are acoustically different. Depending on context [ \ ] is characterised by a short epenthetic vowel-like sound before and/or after the period of tap contact between the tongue tip and the alveolar ridge [2]. And the tap period shows up as a break in waveform and formants. As can be seen in Figure 2 neither spectrogram nor waveform show any such abrupt changes for the dorsal /r/ pronunciation. In a public recognition system these two types of /r/ realisations should ideally also be modelled differently.
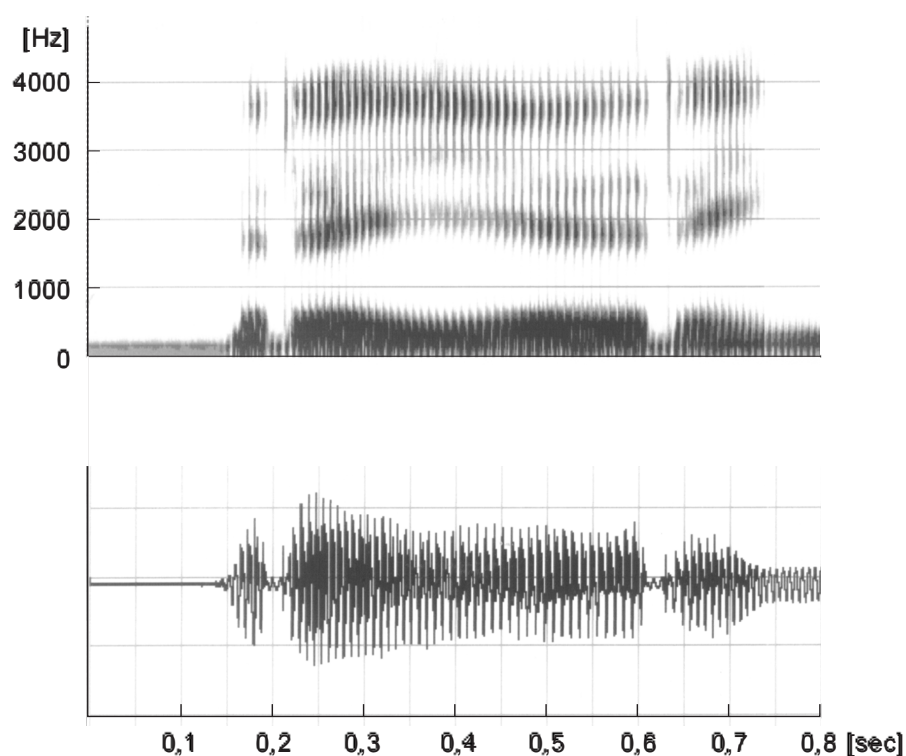
**Figure 1:** Spectrogram, above, and waveform, below, of voiced apical tap pronunciation of /r/ in the Norwegian word "rir" (= (he) rides), [\i:\]. Male speaker.
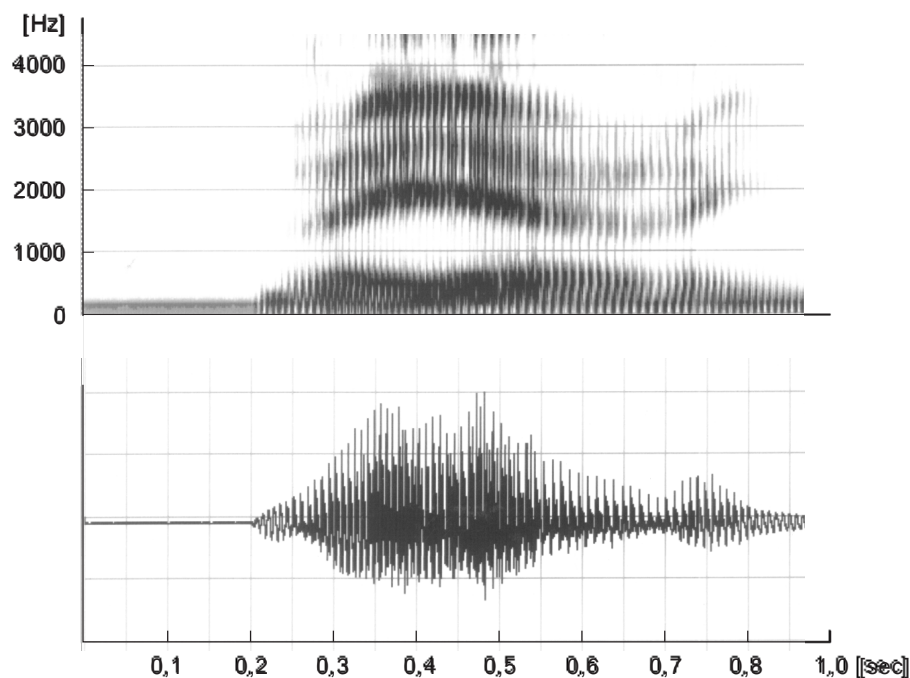


**Figure 2:** Spectrogram, above, and waveform, below, of voiced velar approximant/fricative pronunciation of /r/ in the Norwegian word "rir" (= (he) rides), [ʋi:ʋ]. Male speaker.

# 3. DISTRIBUTION

The distribution of the two types of /r/ pronunciation based on traditional dialect research is shown in Figure 3, with a typically clear cut border line [3]. A thorough auditory analysis of /r/ pronunciation based on the Norwegian part of the European fixed network database, SpeechDat, [4], [5], however, gives a very different picture of present day /r/ pronunciation.
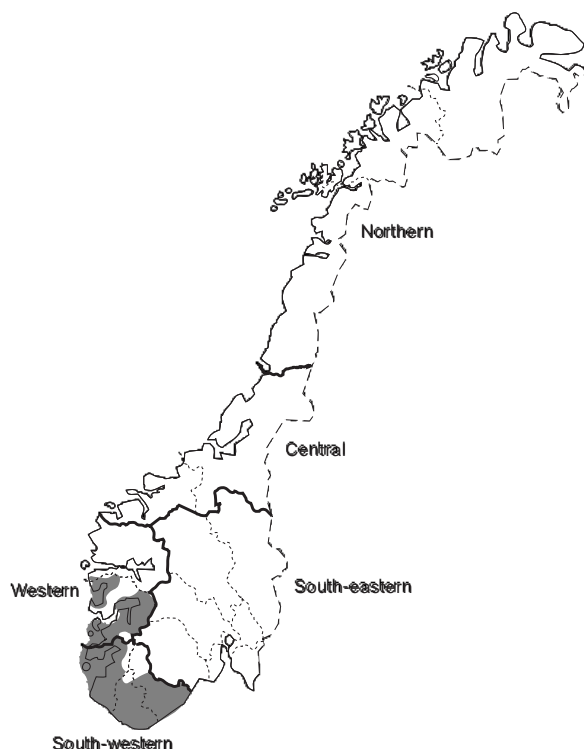


**Figure 3:** Map of Norway showing the division into 5 main dialect areas, Northern, Central, Western, South-western, and South-eastern. Striated areas indicate areas where, according to traditional dialect research, a dorsal /r/ pronunciation is used.

The SpeechDat speakers comprised a sample of 1015 informants (of a total population of 4,4 million) which was representative of the total population in terms of age and dialect. The sexes were equally represented. All except 46 speakers provided information about where they lived and which of 23 dialect regions their dialect belonged to.

The results from the auditory analysis of /r/ pronunciation are shown in Table 1 with the 23 dialect regions pooled into the 5 main regions shown in Figure 3 and informants are divided according to their own assessment of which dialect they speak.

Table 1 shows that [\], alveolar tap, is the most common /r/ realisation for Northern, Central, and South-eastern dialects. The 672 [\]-users constitute 66.2 % of the total number of speakers. If we also include the other two tap-variants, the palatalized tap, [\ʲ], and the velarized tap, [\ᵛ], 769 speakers, 75.8%, of the speakers use some kind of apical tap as their /r/ realisation. In fact, 98.6% of the speakers with a South-eastern dialect use the apical /r/ pronunciation.

The apical approximant, [®], predominates among Lofoten dialect speakers in Northern Norway. The palatalized tap, [\ʲ], is typical of the Oslo area dialect speakers in South-eastern Norway, while the velarized tap, [\ᵛ], is centred to the Molde and Sogn and Fjordane dialects in Western Norway.

The 234 speakers who pronounced /r/ as a velar, [ʌ], or a uvular, [Â], fricative or approximant constitute 23.1% of the total number of informants. The dorsal pronunciation which predominates in the South-western dialects was used by 124 speakers, 93.9%, and in the Western dialects of Norway dorsal pronunciation was used by 78 informants, which was 70.4% of the informants speaking with a dialect from that area.

None of the 23 dialect regions show a uniform /r/ pronunciation. If an /r/ pronunciation map were to be made on the basis of the present auditory analysis, clear cut isogloss lines would have to be subsistuted by broad border areas dividing areas which are predominantly either apical or dorsal /r/. Even more different from the traditional dialect map would an /r/ distribution map be which were based on the region that the call came from.

| Dialect area | Apical /r/ | | | | | Dorsal /r/ | | Sum |
|---|---|---|---|---|---|---|---|---|
| | ® | r | \ʲ | \ | \ᵛ | ʋ | Â | |
| Northern | 7 | 1 | 2 | 101 | 0 | 1 | 0 | **112** |
| Central | 0 | 1 | 2 | 135 | 20 | 0 | 0 | **158** |
| Western | 1 | 2 | 0 | 23 | 11 | 78 | 10 | **125** |
| South-western | 0 | 0 | 1 | 7 | 0 | 121 | 3 | **132** |
| South-eastern | 0 | 0 | 56 | 369 | 1 | 5 | 1 | **432** |
| Non-native speaker | 0 | 0 | 0 | 4 | 1 | 5 | 0 | **10** |
| Dialect unknown | 0 | 0 | 1 | 33 | 2 | 10 | 0 | **46** |
| **Sum** | **8** | **4** | **62** | **672** | **35** | **220** | **14** | **1015** |

**Table 1:** Apical and dorsal /r/ pronunciation of Norwegian SpeechDat informants in different dialect areas subdivided into 5 apical and 2 dorsal variants.

## 4. NORWEGIAN VS. OTHER LANGUAGES

Our analysis of /r/ pronunciation in the Norwegian SpeechDat database shows that traditional dialect maps and surveys may be of limited use in ASR since variation and not uniformity is the rule rather than the exception. To what extent do the Norwegian findings have parallels in other countries? We assume that this will be dependent on two main factors, Firstly, the extent to which there is dialect allegiance vs. a national standard pronunciation and secondly, the extent to which the population is sedentary or mobile.

In Norway, there is no approved standard of pronunciation, and Norwegians, whether they be MPs, teachers, trade union leaders or whatever, tend to use their own dialect in most situations. Consequently, the amount of variation in Norwegian is probably greater than in a country with a widely accepted standard pronunciation. As for mobility, Norway has had a traditional policy of regional development which has supported rural areas, counteracting centralisation to the bigger towns. We would have assumed, therefore, that Norwegians are somewhat less mobile than many other Europeans. However, a forthcoming report shows that this is not the case [6]. In fact, Norway has the highest internal mobility rate of 10 European countries in the survey.

## 5. CONCLUSION

We conclude, therefore, that results from traditional dialect research may be more useful in ASR of other languages than they probably are in Norwegian. Comparisons between databases such as SpeechDat and traditional dialect results will have to be carried out.

## 6. REFERENCES

1. Weijnen, A. et al.(eds.), *Atlas linguarum Europae*. van Gorkum, Assen, 1975.

2. Kvale, K. and Foldvik, A. K., "The multifarious r-sound", *Proc. International Conference on Spoken Language Processing. (ICSLP-92)*, 1259-1262, 1992.

3. Foldvik, A. K., "The pronunciation of r in Norwegian with special reference to the spread of dorsal r." *5th Scandinavian Conference of Linguistics*, 10.5-10. Acta Universitas Lundensis. Sectio 1. Stockholm, 1979.

4. Höge, H. et. al., "European speech databases for telephone applications", *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP-97)*, 1771-1774, 1997.

5. Johansen, F.T., Amdal, I., Kvale, K., "The Norwegian part of SpeechDat: A European Speech Database for Creation of Voice Driven Teleservices", *Proc. Norwegian signal processing symposium (NORSIG-97)*, 40-43, 1997.

6. (Forthcoming): *European Mobility Patterns. CDPO-report*. European Council.