Large-scale corpus based analysis of regional varieties of Danish

Rasmus Puggaard
r.p.hansen@hum.leidenuniv.nl
Leiden University Centre for Linguistics

Yonatan Goldshtein
yonatan@cc.au.dk
Aarhus University

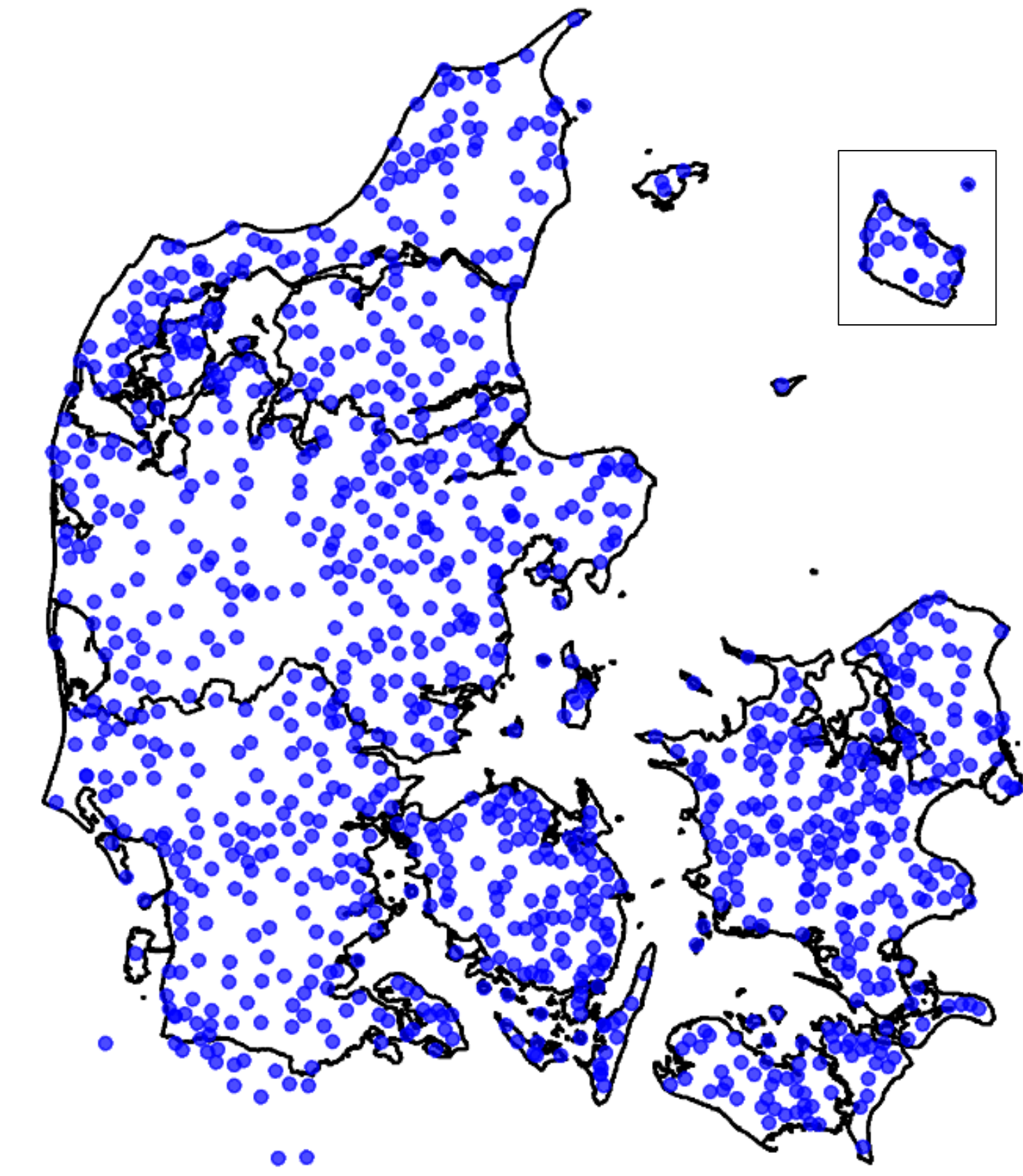Universiteit Leiden

AARHUS UNIVERSITET

Introduction

The **Peter Skautrup Centre for Jutlandic Dialect Research** in Aarhus and the **Department of Dialect Research** (Copenhagen) have been recording dialect speakers for posterity since the 1930s. Taken together, these recordings make up a corpus of stunning size and scope. The recordings have only been used sparingly in the daily work of the centres. Due to technological advancements, a corpus of this scope provides interesting opportunities for e.g. phonetic research.

Life cycle of a recording

Shellac (oldest recordings)

↓

Tape   Digital (newest recordings)

↓

CD

↓

mp3 (Jutlandic collection)   HQ digital restoration (Royal Library)

The full corpus

Sociolinguistic interviews, radio interviews, poetry readings, word lists, conversations, musical performances, etc
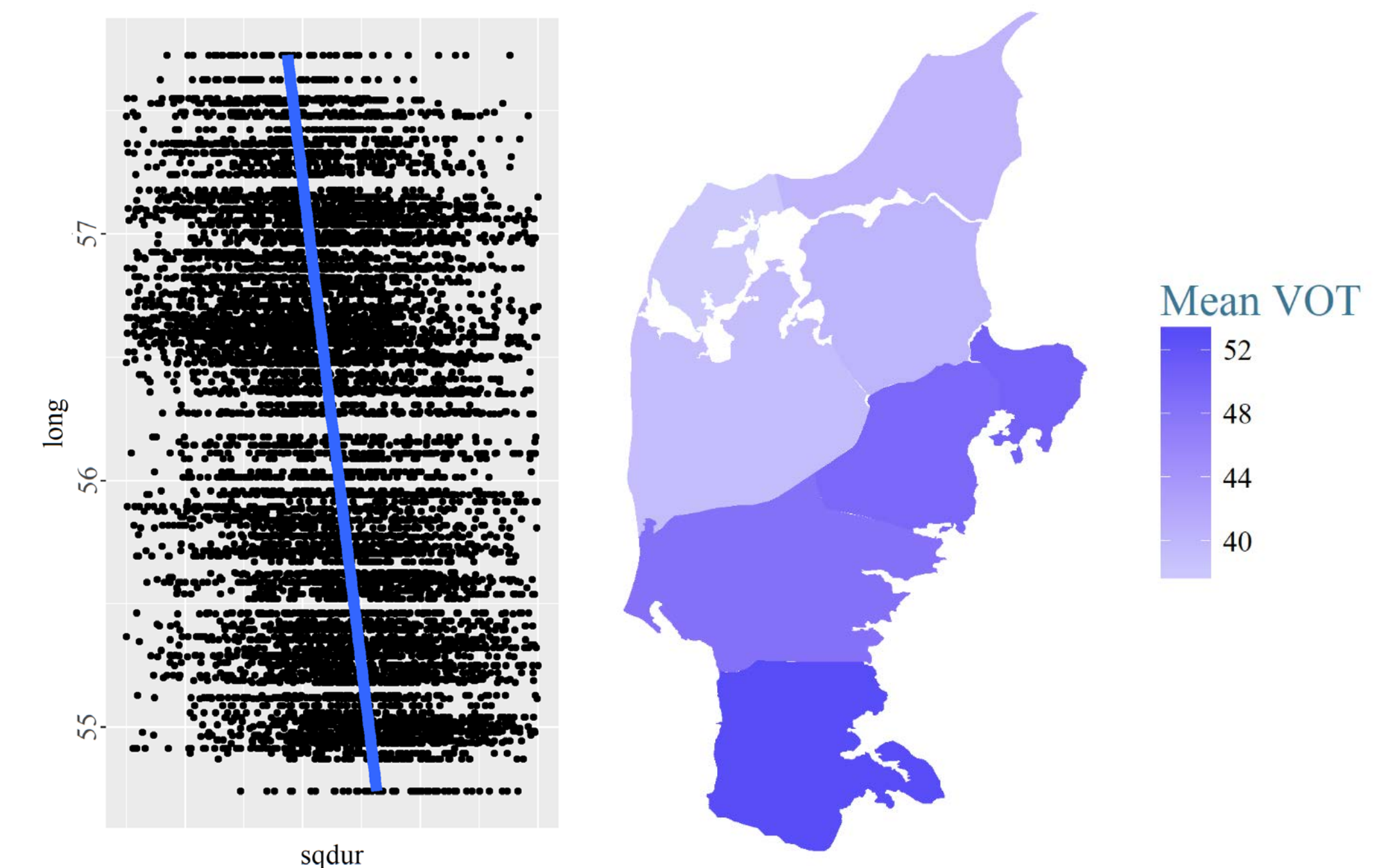
Speakers from **1080 locations** recorded between **1934-2006**. Birth years range from **1849-1986**. Approx. **2090 hours ~ 87 days** of audio.

Audio quality is highly unstable and there are no systematic transcriptions. **But the geographical coverage could hardly be any better.**
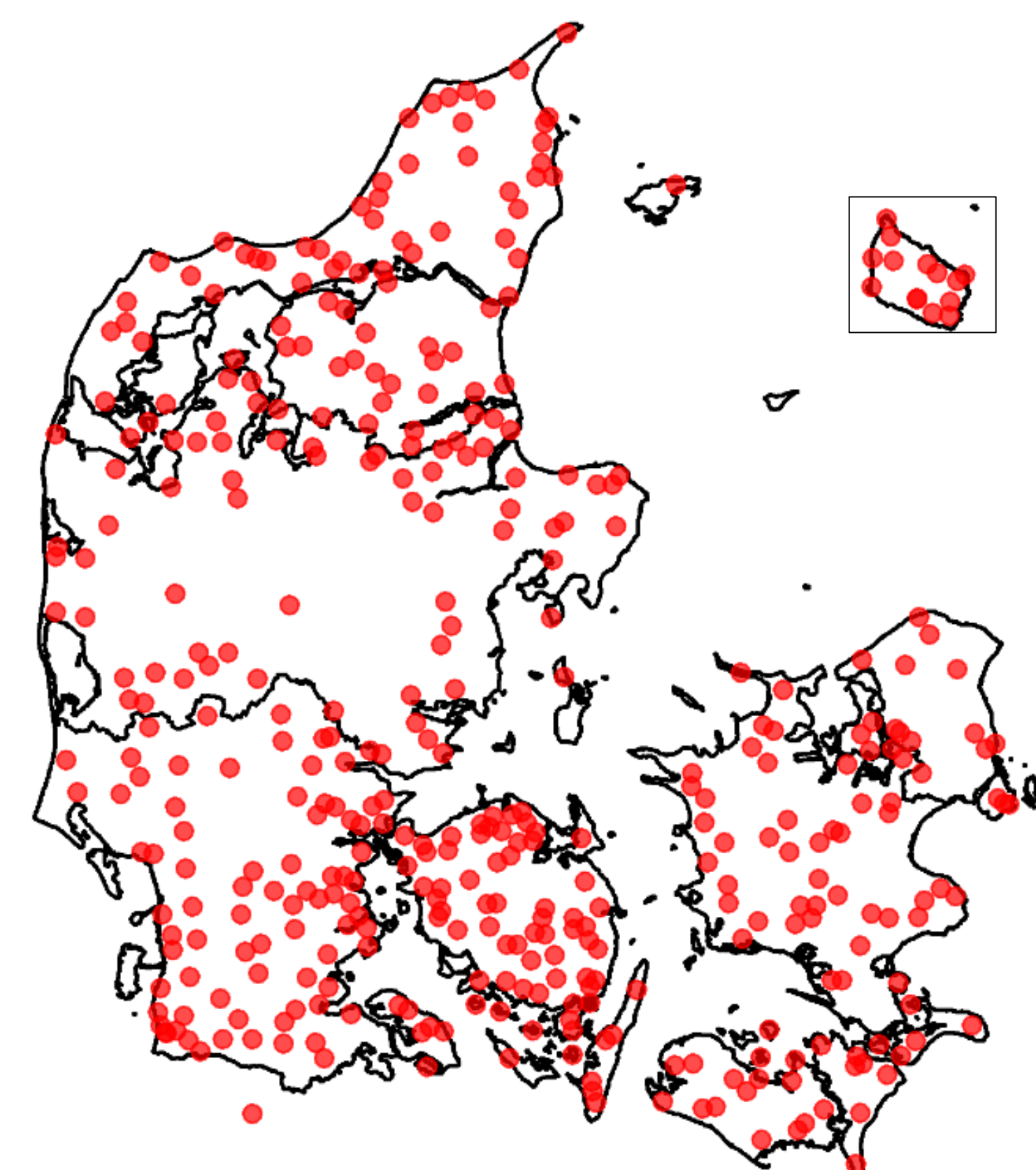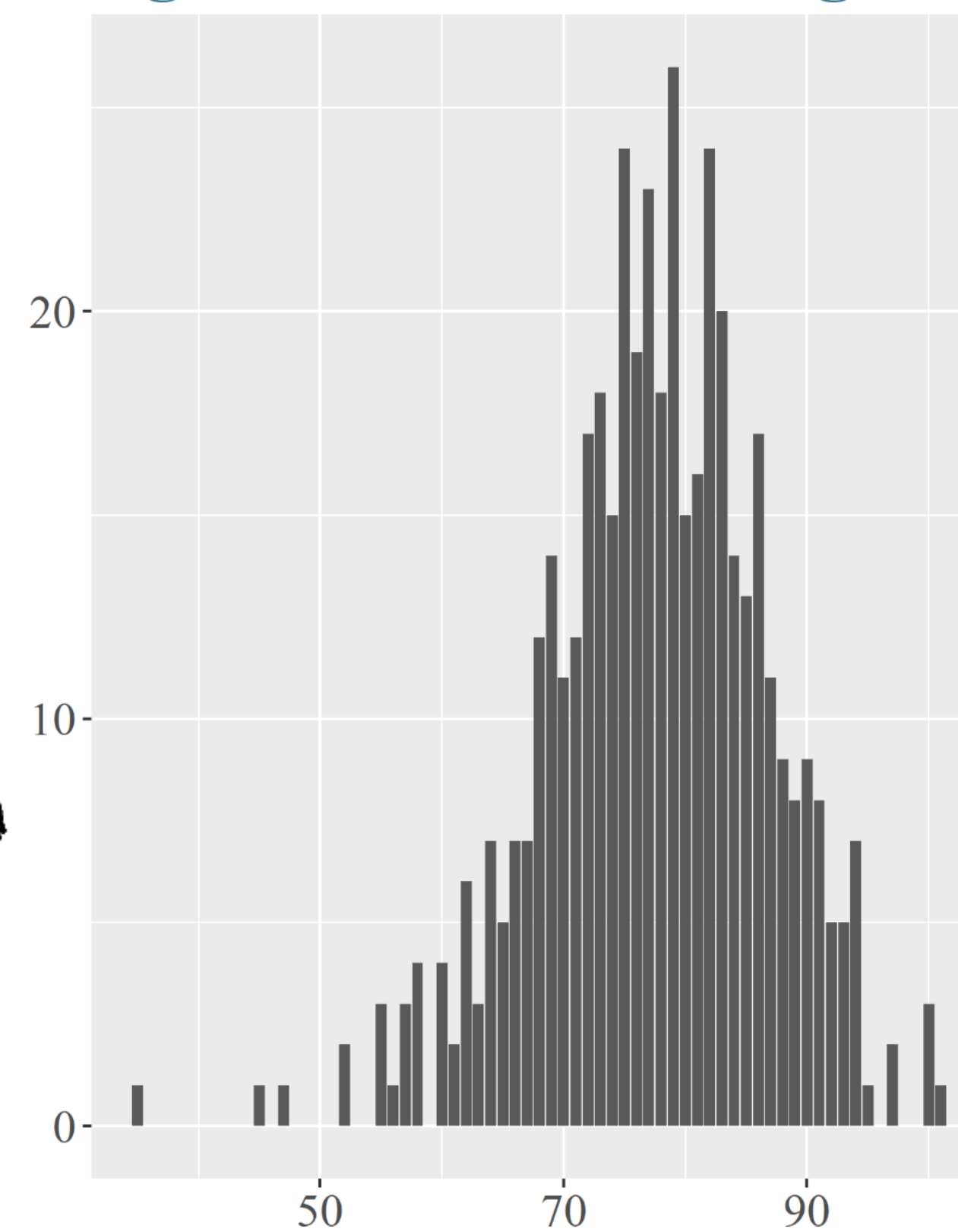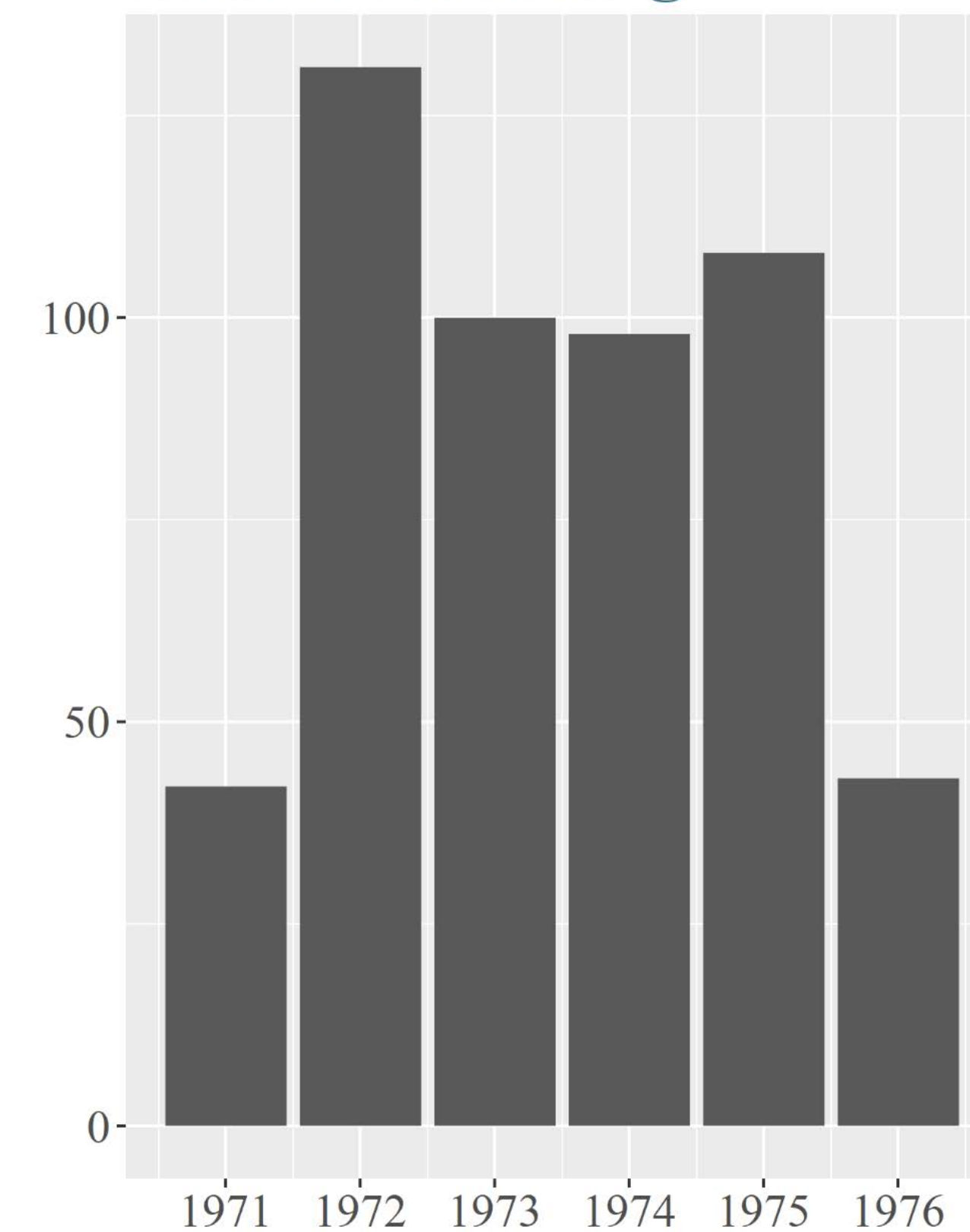

The Royal Library dialect collection

Sociolinguistic interviews, mostly with a single dialect speaker (**68% male, 22% female**). **524 recording sessions** representing **402 parishes**. Approx. **366 hours ~ 15 days of audio**. Relatively stable audio quality. Around 10% of recordings are transcribed. Impressive coverage in spite of gaps.
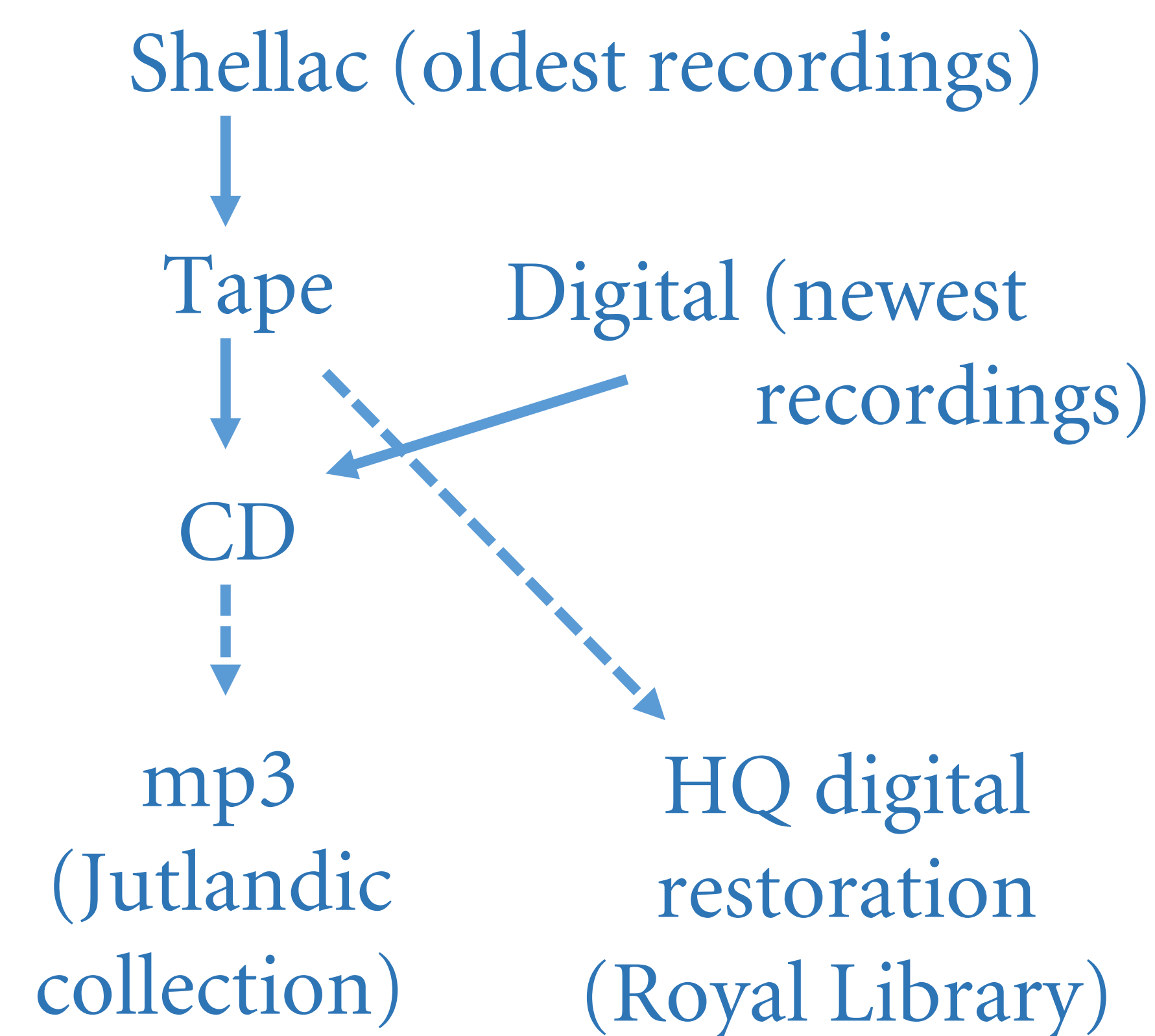

Age at time of recording

Year of recording

Preliminary findings

Preliminary research on **stop realization** shows that **Voice Onset Time is regionally bound** and **decreases from North to South:**

Mean VOT
52
48
44
40

Preliminary research on **the tonal dialects of Southern Jutlandic** shows that the **word accent is typologically affiliated with that of the Peninsular Scandinavian dialects,** but that it might challenge some assumptions of the typology.

^{2}hu:s
CV:S
μ μ μ
H*L HL

hus-PL

h   u   s