


Please cite the Published Version

Drummond, Robert , Dann, Holly, Tasker, Sarah and Ryan, Sadie (2022) The Manchester Voices Accent Van: taking sociolinguistic data collection on the road. *Linguistics Vanguard*, 8 (1). pp. 263-277. ISSN 2199-174X

DOI: <https://doi.org/10.1515/lingvan-2022-0050>

Publisher: Walter de Gruyter

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/630189/>

Document DOI: <https://doi.org/10.23634/MMU.00630189.00303645>

Usage rights:  [Creative Commons: Attribution 4.0](https://creativecommons.org/licenses/by/4.0/)

Additional Information: This is an Author Accepted Manuscript of an article that has been published in final form in *Linguistics Vanguard* by Walter de Gruyter publishing. This Author Accepted Manuscript has been made available with no embargo and a CC-BY-4.0 licence with permission of the publisher to meet the requirements of the UKRI Open Access Policy.

Data Access Statement: Due to the nature of data processing, the data will be made openly available in a timely manner following publication. Details of the data availability are available from Manchester Metropolitan University's research repository at <https://doi.org/10.23634/MMU.00630878>

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

This is the accepted version of the article. The published version is available in *Linguistics Vanguard*: <https://doi.org/10.1515/lingvan-2022-0050>

Title

The Manchester Voices Accent Van: taking sociolinguistic data collection on the road

Abstract

This article details the methodology behind the Manchester Voices Accent Van, and the accompanying online Virtual Van. In 2021, the project travelled around Greater Manchester in a van converted into a mobile recording booth, asking people to climb aboard and take part in an unsupervised interview about language and identity in the region. Participants could also take part from their own home through a bespoke website, called the Virtual Van, which asked the same interview questions as the physical van and recorded speakers through their computer/phone microphone. With a view to informing others who might want to use similar methods in the future, we present a detailed description of the methodology here, as well as an overview and sample of the data collected. We conclude with a reflection on the elements of the data collection that went well, and a discussion of improvements and considerations for future research using this methodology.

Keywords: sociolinguistics, language variation, data collection, remote data collection, community.

1. The project

Manchester Voices is a large-scale sociolinguistics project exploring the accents, dialects and identities of people in Greater Manchester, a city-region in northwest England with a population of almost 3 million. Greater Manchester was created in 1974, and comprises ten boroughs, most of which previously belonged to the historic counties of Lancashire, Cheshire, Derbyshire and Yorkshire. As a result, and given the established linguistic diversity in the region (e.g., Trudgill 1990: 63), there are some interesting stories and experiences to uncover regarding local and supra-local identities and their relationship to language. The project overall has a community-oriented, folk-linguistic (Niedzielski and Preston 2003), citizen sociolinguistics (Rymes and Leone 2014) approach. Key strands include: 1) perceptual dialect maps using an innovative online version of the *draw-a-map* task; 2) analysis of historical recordings from the area (Ryan et al. 2022); and 3) poetry workshops exploring language and identity. The fourth strand, described in this paper, is the Accent Van and the accompanying Virtual Van, the part of the project tasked with collecting spoken data from people living across the ten boroughs of Greater Manchester (Figure 1).

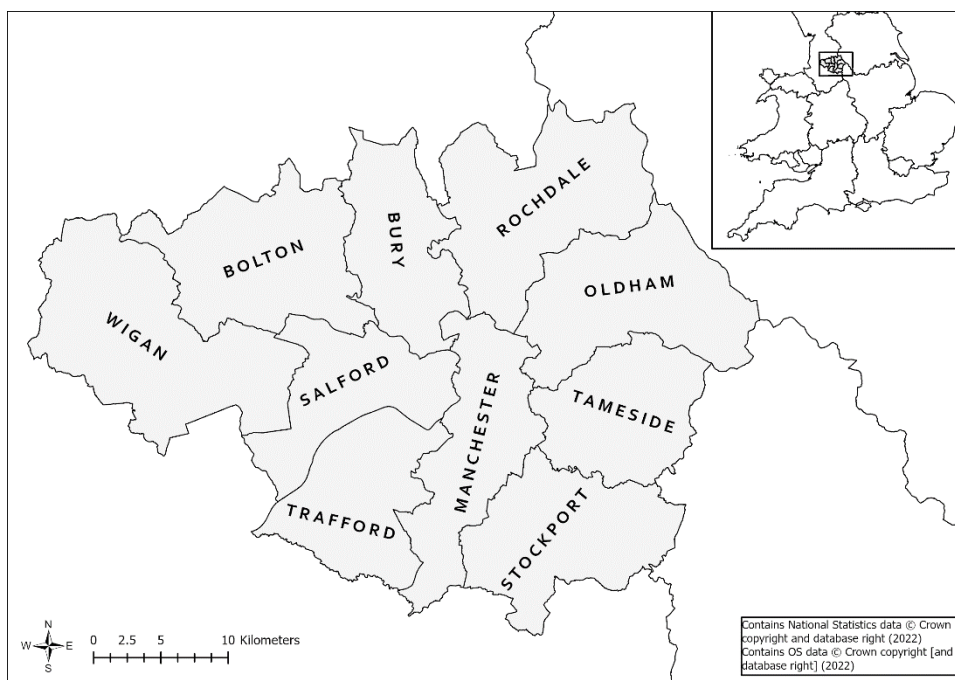


Figure 1: Map showing the ten boroughs of greater Manchester, and the location of Greater Manchester within the UK.

There are two planned types of outputs for the project: those aimed at an academic audience, including journal articles and conference papers, and equally importantly, a permanent screen-based interactive installation at Manchester Central Library. This will be open to the public, and will allow visitors to explore the data, analysis, and interpretations from all strands of the project, including video and audio clips recorded in the Accent Van and Virtual Van.

2. The Accent Van

The Accent Van was conceived as an innovative way of collecting spoken data from a wide range of people, some of whom might not otherwise take part in sociolinguistic research. Traditionally, such research is carried out in the form of sociolinguistic interviews – pre-planned events where a researcher and participant meet and record a semi-structured interview, with the general aim of eliciting natural and less guarded speech than one might expect in a formal interview situation (Labov 1972). The interviews tend to cover a wide range of topics, and while they can include

specific questions on language use, often the interviewer deliberately tries to draw attention away from such linguistic self-reflection. However, there is also value in a complementary type of sociolinguistic data – that which explicitly explores people’s thoughts on their own, and others’, ways of speaking. While arguably not as natural as the language recorded through more traditional methods, it quickly opens the door to attitudes and ideologies that might be hard to explore otherwise.

The Accent Van is an attempt to gather this kind of reflective data, and to provide an accessible method of spoken data collection for linguistic analysis. It follows recent projects which have sought to explore both more innovative methods of sociolinguistic data collection, and ways to make sociolinguistic research a public resource. For example, Sneller et al. (2022) and Hall-Lew et al. (2022) discuss the use of self-recorded diaries (audio and audio-visual respectively) as both sociolinguistic research data, and as a method of recording and exploring different (often marginalised) groups’ experiences of Covid-19. Similarly, the QuakeBox Corpus (Walsh et al. 2013; Clark et al. 2016) consisted of audio and video recordings made in a transportable recording studio in a shipping container. Travelling around different locations in Canterbury, New Zealand, the QuakeBox recorded 722 people’s stories of the 2010-2011 earthquakes. The recordings now exist as a sociolinguistic resource, and an historical record of the time.

Essentially, the Van is a mobile recording studio. In this particular context, this was a Volkswagen Caravelle Executive Vehicle kitted out with a video camera and tripod (Sony Handycam HDR-CX405 Camcorder and a Hama Star 75 tripod), a voice recorder (Zoom H2n) with a built-in microphone, a studio light (Neewer 176 LED Video Light), and a tablet (Samsung Galaxy Tab A). Over the course of 3.5 months in 2021, the Accent Van travelled around the ten boroughs of Greater Manchester visiting live events, parks, libraries, community centres, mosques, schools, and colleges with a view to inviting people to climb aboard and reflect on the way they speak. In addition to the physical Van, we also created a Virtual Van experience that people could access online. The rationale behind and practicalities of the Virtual Van are discussed in Section 3.

2.1 Decisions setting up

The particular model of the Van (Figure 2) was selected for reasons of practicality and comfort. From a practical perspective it was important that there was a table on which to place the tablet and voice recorder, and that there was room for up to three participants to sit comfortably together on the back seat. From a comfort perspective it was important that the van was clean, professional looking, and inviting, preferably with tinted windows and/or sun blinds to allow a degree of privacy. We were able to lease a new vehicle, which we branded using professionally designed vinyl vehicle graphics. The branding was important, both as a way of advertising the project, but also as a way of reassuring prospective participants of the authenticity of the research by including the name and logos of the university and the funding body.



Figure 2: The Accent Van

Audio recordings were made in .wav format using a 44.1 kHz sampling rate with 16-bit precision, and video recordings were made in .mp4 format. Both were saved onto an SD memory card then transferred onto a PC. The reason for including video recording was primarily to provide additional content for the public-facing final installation at Manchester Central Library that would be more engaging for people to interact with than audio clips alone. In addition, as the project data will ultimately be shareable, video recordings will allow future research into the use of gesture as part of the communicative process. However, this is not a neutral decision – the use of video potentially adds an additional layer of complexity regarding how willing people are to take part, and how comfortable they will feel when speaking. This will be addressed in section 6.

The tablet was put into a case which had its own stand, enabling it to be placed on the table in front of the participants (see Figure 3). It ran a simple PowerPoint slideshow containing audio. Each slide displayed a question which was simultaneously shown on the tablet screen and played through a Bluetooth speaker placed in the rear of the vehicle. Participants navigated through the questions by tapping on an arrow labelled “Next”



Figure 3: Inside the Accent Van

One crucial decision that had to be made early in the process of setting up was what voice to use for the audio version of the questions. One of the reasons for using a virtual interviewer was to ensure consistency and to minimise interviewer effects. However, we it should be noted that different demographics may still respond to a virtual voice differently (see, e.g., Zellou et al. 2021). There is also potentially something quite distancing about using technology rather than a real person, so to minimise this, we used recordings of a real person rather than rely on a computer-generated voice. We also decided to avoid a typically ‘standard’ accent such as RP or Standard Southern British English (SSBE) for the same reasons. Instead, we used an ‘insider’ voice and advertised for a voice artist among the university student population. Clearly this decision is not without issue, as choosing a voice from one particular area of Greater Manchester might also affect how the interviewer is perceived, and possibly be alienating for participants who recognise the voice as different to their own in some way. However, we felt that overall it would still be preferable to use a local voice over an RP or SSBE voice, which we felt could have created unnecessary distance between the interviewer and the interviewee, and make participants less likely to speak freely. Eventually, we selected the voice that we simply felt was the most friendly, a young woman from Bolton, who had a fairly typical accent for someone from the local Asian community. We worked with her to record the interview questions, instructions, and a short welcome and goodbye message.

2.2 Selecting venues

Selecting locations and events to take the Accent Van to was based around the following considerations. Firstly, we aimed to visit each borough at least once. Secondly, it had to be a location

where we could expect a fair amount of passing foot-traffic, or where people would naturally be gathered. Thirdly, it had to be an event where we had a good chance of meeting people from particular demographics in order to maintain a broadly accurate representation of the diversity within the region. This was not done systematically in terms of quotas, as we didn't want to exclude anybody who wanted to take part. However, we did try to target particular community events, such as a community barbecue at a local mosque and a public park with holiday activities for children, in order to increase our chances of engaging with people of specific ages, ethnicities, or other social groups.

Accent Van stops were always arranged in advance with whoever had oversight of the venue or event. Sometimes this was necessary to gain access to a particular event (e.g., a music festival), and at other times it allowed us additional access that we would otherwise not have had (e.g., taking the Van into a pedestrianised area of a public park). Several times it gave the organisers of an event the opportunity to create additional publicity for their event by advertising our presence. On a few rare occasions, a venue we targeted proved to be impossible to secure, either due to the unavailability of our own staff at relevant times, or due to seemingly impenetrable layers of bureaucracy and required permissions. However, in the main, people seemed very keen to welcome us.

2.3 Staffing the Accent Van

For most scheduled stops the Van was staffed by two or three people, usually the first author to drive and set up the equipment, and one or two student research assistants. The role of the assistants was to encourage people to take part, to explain the project, and to go through the consent information. Everyone working on the Van wore clothes visibly branded with the project name and logo. Each member of the team was given a t-shirt and hooded sweatshirt which they were expected to wear on each outing, and the Van itself was equipped with three waterproof jackets and two umbrellas for shared use.

2.4 The data collection process

Each location or event was different when it came to inviting people to take part. On some occasions, the visible presence of the van, along with a pavement board describing the key points of the research, were enough to prompt people to volunteer to take part. On other occasions, more active recruitment was required. Often, once one person had taken part, this encouraged others to do the same. Working in our favour was that people generally have something to say about, and enjoy talking about, accents and dialects. Before, or at the point of, entering the van, participants were provided with a detailed information sheet, and the project was verbally explained to them. They were also taken through the consent form, which allowed them to choose whether they wanted to be audio and video recorded or only audio recorded, and whether they were happy for their recordings to be used in the public-facing installation. For under 16s, we asked for their assent to the same questions, as well as the consent of a parent or carer. We also recorded demographic information in the form of age group, gender, ethnicity, and the first part of their postcode. The person would then sit on the back seat of the Van, facing the camera and the table with the tablet and audio recorder. One of us would then start the recording devices, leave the Van, and close the door. It was felt that having no researchers in the van during the interview allowed for a more private, reflective space, that might in some ways mitigate the presence of a camera. The interviews took anywhere from 5 to 45 minutes, depending on how engaged the person was.

Participants could also take part in groups of two or three people at a time. When people were clearly part of a pair or group, we encouraged them to take part together rather than individually. This happened on 28 occasions, with 26 pairs and two groups of three. There is no doubt that the

inclusion of more than one person at a time affected the nature of the recording. This will be addressed in section 6.

3. The Virtual Van

The Virtual Van aimed to mirror the physical van experience as closely as possible and was set up for three reasons. Firstly, it enabled us to continue the project during the Covid-19 pandemic. Although we were hopeful that the physical Van would be able to go out, a virtual option was a contingency plan. The pandemic also made it likely that certain groups of people were more likely to stay away from public events. Secondly, it allowed for greater accessibility to the project, as the majority of people (in the Western world) at least have access to a smartphone (see Leemann 2021). Although every reasonable effort was made to make it easy for anybody to take part in the physical van experience, the van did not have disabled access. Thirdly, if people were interested in taking part but were not available when we were in their area, it allowed them to take part in their own time.

The Virtual Van was a browser-based web-app which was designed specifically for this project by externally hired software developers. Although expensive, this approach ensured a user-friendly platform that could record for as long as the participant chose, circumventing some of the issues raised in Sneller et al. (2022) around long buffering times. The app existed on a subdomain of the project website, and worked on most devices, including tablets and smartphones. It began by taking the same demographic and consent information as the physical Van. The site then asked for permission to access the device's microphone, and the questions were presented in the same way as the physical Van. The recording started automatically and was uploaded automatically to the software developer's servers when the participant had finished and clicked "submit". The Virtual Van collected audio recordings only. Regarding audio quality, a small selection of recordings had to be discarded due to audio distortions, although it was unclear whether this was due to the participants' recording devices or an issue with the web-app. As indicated by Leemann (2021), we found that all remaining self-recordings were usable for, at minimum, the acoustic analysis of vocalic variation. Further examination would be required to explore the suitability of all these recordings for the analysis of other acoustic parameters, such as centre of gravity.

4. The "interview"

The interview consisted of 14 questions and an optional structured elicitation task. It touched on a range of topics around the themes of language, place, and identity. For example, participants were asked for their opinions on the concept of Greater Manchester and about their local area, their thoughts about their own speech and how they feel it is perceived by others, and for linguistic examples that were typical of where they live. Additionally, they were asked some basic biographical questions to find out how long they had lived in different areas of Greater Manchester. The full list of the questions asked is provided in the appendix.

[associated audio-1-drummond with Figure 4]

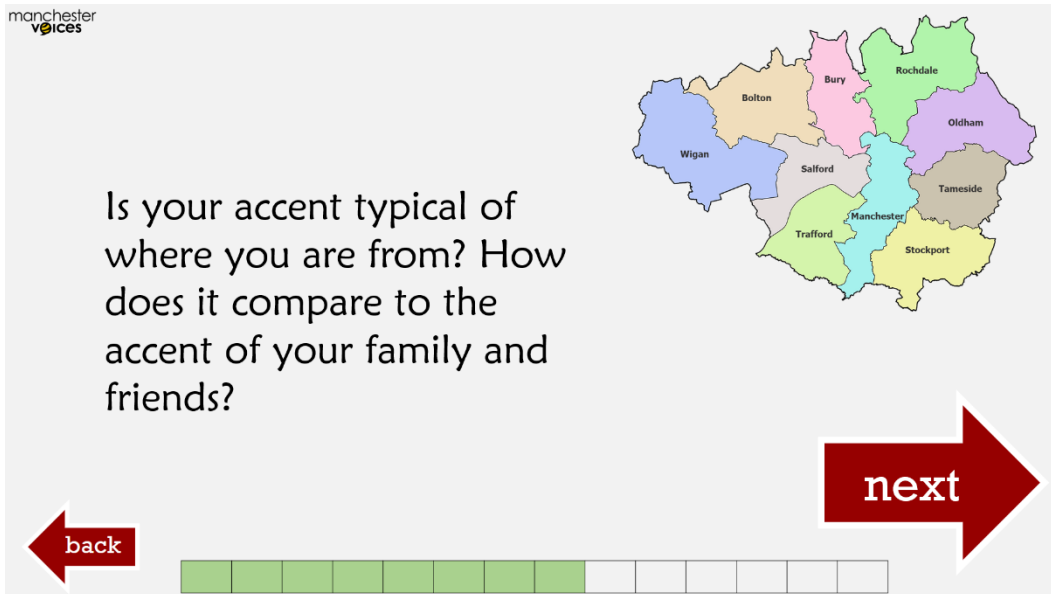


Figure 4: The Accent Van questions, presented on a tablet

Adult participants were given all 14 questions in the same order, and they could see how far along they were by looking at the progress bar at the bottom of the screen (Figure 4). For participants under the age of 16, some questions were removed, where it was felt that children may lack the knowledge and awareness to be able to respond. Full details of this can be found in the appendix.

At the end of the interview, participants were then given the option to “Finish”, or to complete a map task (Anderson et al. 1991), where they were asked to describe a route along a dotted line to an imaginary other person who did not know the route, with reference to the items that they came across along the way (Figure 5). This task was intended to elicit tokens of accent features that vary across Greater Manchester in a consistent set of linguistic environments, some of which (such as the PALM vowel) occur relatively infrequently in casual speech. A map task was chosen in place of more traditional elicitation tasks, such as a word list or reading passage, as they have been found to be more comparable to interview speech (Boyd et al. 2015), and have been employed successfully in variationist research such as Cardoso (2015) and Dann (2019).

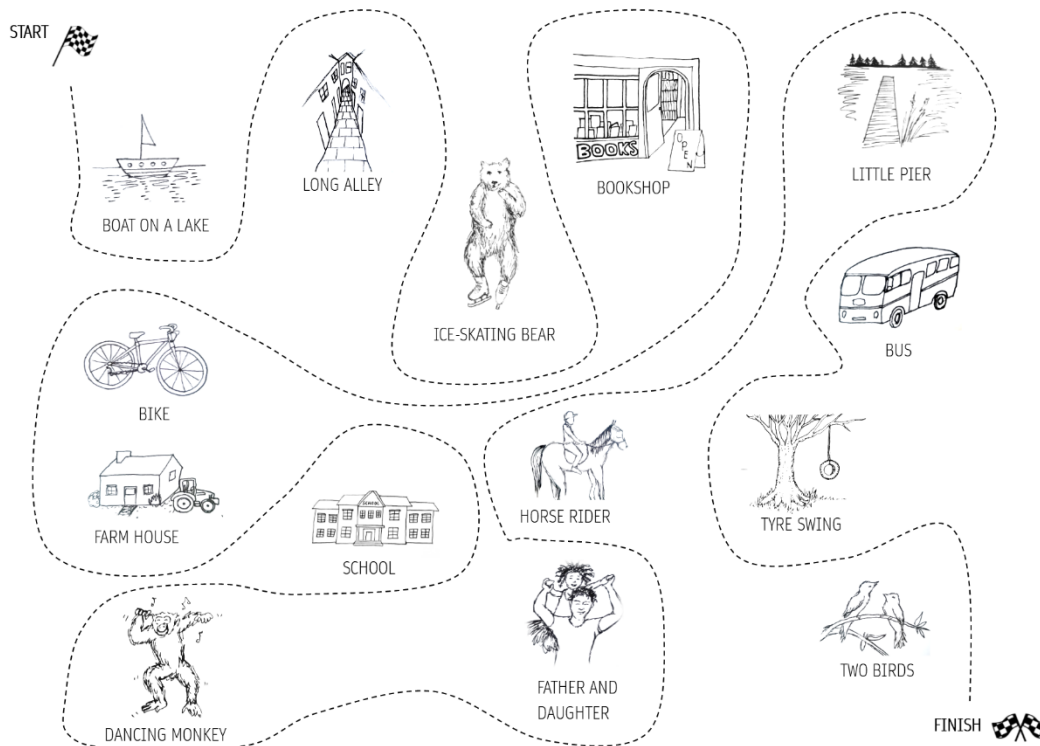


Figure 5: The Map Task

5. The data

In total, 421 participants were interviewed – 200 from the Accent Van and 221 from the Virtual Van. Prior to analysis, 25 of the Virtual Van speakers were removed due to either not ticking all the required consent checkboxes or because of audio distortions. This is clearly a risk with any unsupervised data collection method, in that the researcher relinquishes control of aspects of the process, meaning that some loss of participants is inevitable. Additionally, just one Accent Van speaker was removed due to not living in Greater Manchester.

Table 1 provides an overview of the remaining data, and Figure 6 shows the locations where the Accent Van stopped and the locations of the participants within Greater Manchester. This demonstrates that this methodology was effectively used to collect a large volume of data, relatively well balanced across areas of the region, ages, and genders.

Table 1: An overview of the Accent Van and Virtual Van data

Total participants	395
Accent Van participants	199
Virtual Van participants	196
Total Accent Van stops	30
Average length of stop	4h
Median participants per stop	6
Proportion agreed to video	170/196
Total length of all interviews	72h 52m 12s
Maximum/Minimum/Average interview length	45m 38s / 2m 44s / 11m 22s
<i>Participant Demographics</i>	
Grew up in Greater Manchester	310
Lived in one borough of Greater Manchester most of their life	200

<i>Age Group</i>	
Under 18	47
18-25	29
26-45	109
46-65	143
66+	67
<i>Gender</i>	
Male	177
Female	217
Non-binary	1
<i>Borough</i>	
Bolton	28
Bury	48
Manchester	81
Oldham	35
Rochdale	39
Salford	27
Stockport	37
Tameside	28
Trafford	26
Wigan	46

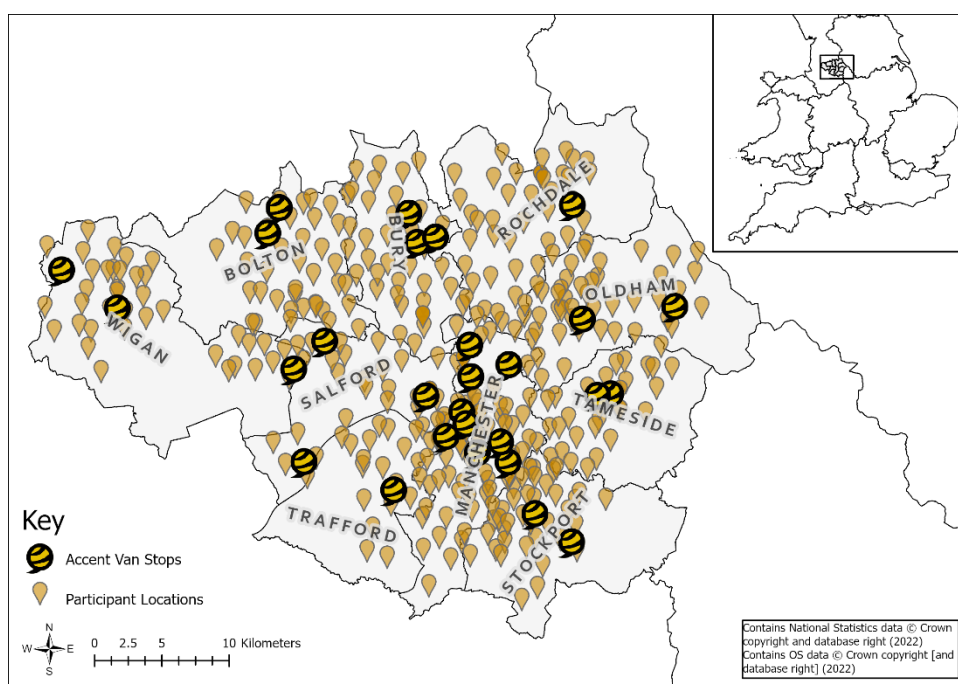


Figure 6: Maps of Greater Manchester showing the stops made by the Accent Van and the approximate residential locations of the participants from both the Accent Van and the Virtual Van.

The data allows for two complementary types of analysis. Firstly, the quantitative analysis of accent and grammatical features of speakers from across the entire region of Greater Manchester is possible. The data is stored in a searchable corpus using the software LaBB-CAT (Fromont and Hay 2012), allowing it to be explored and analysed in various ways. It should be acknowledged that given

the relatively short interviews compared to more traditional variationist research, the number of tokens per type per speaker will be low for some participants. This necessarily makes analysis of inter-speaker variation at an individual level (and intra-speaker variation) more of a challenge, depending on the frequency of the variable of interest. However, this should be balanced against the opportunities that are provided by having data from such a wide range of speakers, meaning that the data is particularly suitable for more macro-level analyses, such as according to speaker age, ethnicity, gender, or location. The ability to gather speech data from a large number of participants across a wide geographical area is a particular advantage of this method of data collection. It is therefore an especially useful method for examining linguistic variables which vary geographically within a region. The quantitative analysis is currently in progress and will be reported in future publications.

Secondly, the type of data collected also allows for qualitative analysis of participants' responses. Participants were asked to reflect on their own and others' language, as well to share their thoughts about their own and other areas within Greater Manchester. This direct approach had the benefit of eliciting rich and varied, and often very reflective responses from participants across a range of topics, as shown in the example clips below. Qualitative and content analysis of such responses could provide insight into, for example, questions relating to local identity and orientation, language attitudes and awareness, and folk-linguistics. In addition, combining such analyses with linguistics analyses can add another layer of understanding of how and why speakers may employ certain linguistic features.

[\[associated video-1-drummond\]](#)

[\[associated video-2-drummond\]](#)

[\[associated video-3-drummond\]](#)

[\[associated video-4-drummond\]](#)

[\[associated video-5-drummond\]](#)

[\[associated video-6-drummond\]](#)

[\[associated video-7-drummond\]](#)

Clips such as these also provide an important way of engaging the public with the outcome of the project, through sharing on social media, and in forming a part of an interactive installation. When the recording was audio only, either through the choice of the participant in the Accent Van, or through the recording being made in the Virtual Van, we added 'B-roll' footage of the relevant borough to make the clip visually appealing for the installation.

6. Reflections

Overall, the Accent Van and Virtual Van were very effective tools for collecting sociolinguistic data, and the approach demonstrated the spirit of the research as a community-based project. The Van enabled the researchers to not just attend but take part in local community events, in the sense that we were often advertised as one of the activities that visitors could engage with as part of their visit. In addition, the style of questions and unsupervised nature of the interviews centred local community knowledge and opinions, which is at the heart of the public-facing aspect of the project.

In this way, the project was particularly successful, as we have been able to create a public resource with almost 400 interview clips representing a snapshot of anecdotes, thoughts, and attitudes of a cross-section of Greater Manchester society. This will be a place where the members of the public can both access the results of our research in an accessible format and explore and reflect on issues around language and identity in their local community.

The unusual and innovative methodology of the Accent Van also captured the imagination of both the local community and media. The Accent Van tour gained a great deal of media attention, both local and national, which played a key role in attracting participants and helped us in booking events to attend. The success of the Virtual Van was due in large part to media coverage of the Accent Van directing interested parties to the website. We suspect that if the Virtual Van had existed alone, it would have needed significantly more promotion and targeted participant recruitment.

We learnt some lessons throughout the process which would affect how we approached future research of this kind. Firstly, there were occasionally some issues with navigating the questions on the tablet, particularly for older participants. In these cases, we had to intervene, which broke the sense of privacy as the door had to remain open or be repeatedly opened and closed. Clearly, familiarity with touchscreen technology should not be assumed. However, as the tablet is ideal in so many other ways, we would suggest adding a physical button to the setup, connected to the tablet via bluetooth.

Although at times we struggled for participants, at others we had more than we could cope with, as there was a 15–20-minute wait between each interview. Although we did use the time to go through the consent information, there was still a degree of waiting around. At times we felt a little bit like salespeople, trying to catch the attention of passers-by who were clearly wary of being cornered into buying something or taking part in a survey, and we would find people speedily walking past us avoiding eye contact. We mitigated this by introducing a pavement board – a large sign which gave the broad details of the project – part way through the Accent Van tour (Figure 7). This helped a great deal as it gave potential participants the opportunity to quietly inform themselves what was happening before engaging with us. It also gave us a ‘way in’ to a conversation with people if we saw them reading the board.



Figure 7: The Accent Van outside a local library

The presence of a video camera is clearly going to be daunting for some people, yet relatively few people (26 out of 196) opted to have it switched off when given the choice during the consent process. It is impossible for us to measure any effect the camera might have had on an individual's speech in relation to the observer's paradox (Labov 1972), but we suggest that the privacy of the situation and the lack of an interviewer goes some way to balancing out any feelings of discomfort. Crucially, given that the situation was the same for all participants, there is consistency across the sample in this regard.

The effect of people taking part in pairs or in groups of three was largely positive. Having someone else there allowed for a discussion between the participants, with people reacting to each other's responses to the questions. On the other hand, several participants were quite talkative by themselves, and it might have been that they would have felt pressured to speak less if someone else were there.

On some occasions background noise from outside the van was a potential issue, particularly at public events, such as music festivals. Although we parked the Van in as quiet a location as we could while still being visible and accessible, we ultimately decided that background noise was an unavoidable aspect of the data. However, these recordings were still clear enough to be analysed acoustically, and a degree of background noise added to the authenticity of the community aspect of the research, especially with a view to it being used in the public installation. Figure 8 shows a representative example of the quality of the audio at one of the loudest locations we recorded in (a music festival). Despite the visible background noise, the amplitude of the speech is generally much higher, so for higher amplitude parts of the spectrogram, such as formants, this noise was not problematic for phonetic measurement. However, such background noise is less ideal when taking measurements at higher frequencies or in analysis of lower amplitude sounds, such as fricatives. For example, in Figure 8, it is relatively hard to distinguish the spectral peaks of /f/ from the background noise. This does not mean that the data cannot be used for the analysis of such sounds, but that

extra care is needed in speaker selection and in checking boundaries and measurements for these sounds.

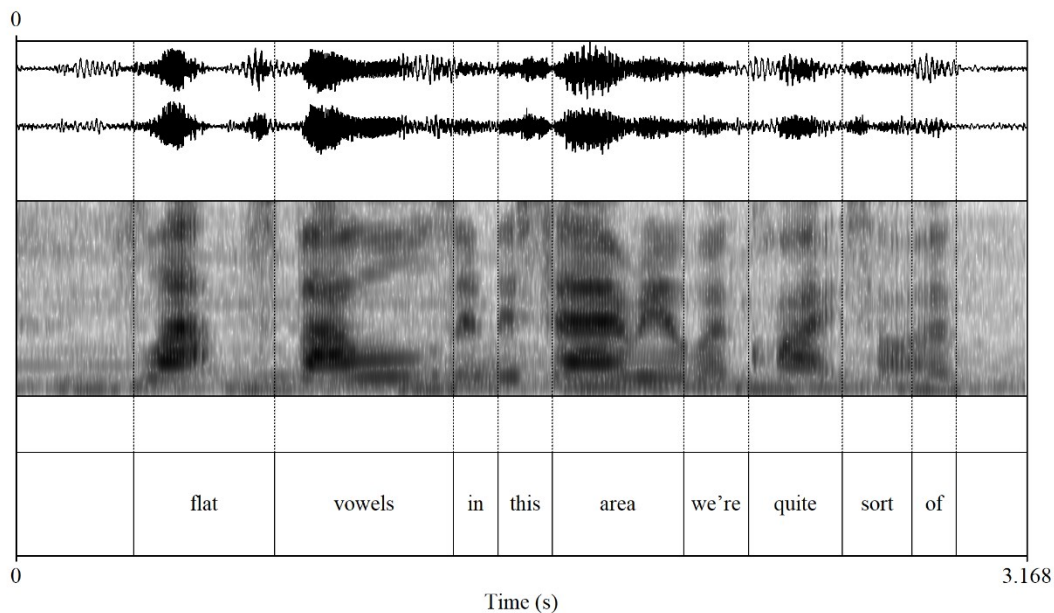


Figure 8: A waveform and spectrogram from a recording where there was noticeable background noise.

Finally, it is also important to note the high costs associated with both the Accent Van and the Virtual Van. Van hire, equipment and staffing were all costly. This was particularly frustrating on the occasions where we spent an entire morning or afternoon at a stop with three staff, only to recruit one or two participants. The Virtual Van was made specifically for the project, ensuring a slick user experience and reducing the need for the participant to have much technical knowledge in order to take part online. As noted by, for example, Leemann et al. (2020), the need for technical knowledge is a drawback of much online research and, as such, we believe that without the funds to create the bespoke website, this part of the data collection would have been much less accessible.

7. Summary

This paper has presented an overview of the methodology behind the Accent Van and its virtual counterpart, which was used to collect sociolinguistic data for the Manchester Voices project. We have detailed some practical details and considerations, with the hope that this will help to pave the way for future research using similar methodologies. In addition, in line with other community-oriented sociolinguistic projects such as Hall-Lew et al.'s (2022) Lothian Diary Project, Sneller et al.'s (2022) MI Diaries Project, and the QuakeBox corpus (Walsh et al. 2013), we have explored the benefits of looking beyond the traditional sociolinguistic interview to consider the value of the recordings and reflections not just as a site for linguistic analysis, but as a public resource.

Acknowledgements

We would like to thank Erin Carrie for her work on the Manchester Voices project in the early stages. We are also extremely grateful for the supportive and constructive comments made by two anonymous reviewers, whose suggestions have made this a far stronger article. Finally, we would like to thank all the people of Greater Manchester who took the time to share their voices, opinions and stories in The Accent Van or the Virtual Van in 2021 and 2022. Without them, the project wouldn't exist.

This research is part of the Manchester Voices project, which was generously funded by the Arts and Humanities Research Council. Project Grant: AH/S006125/1.

Due to the nature of data processing, the data will be made openly available in a timely manner following publication. Details of the data availability are available from Manchester Metropolitan University's research repository at <https://doi.org/10.23634/MMU.00630878>

References

- Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister & Jim Miller. 1991. The HCRC map task corpus. *Language and Speech* 34(4). 351–366.
- Boyd, Zac, Zuzana Elliott, Josef Fruehwald, Lauren Hall-Lew & Daniel Lawrence. 2015. An Evaluation of Sociolinguistic Elicitation Methods. In *Proceedings of ICPHS 2015*. Glasgow.
- Cardoso, Amanda Beth. 2015. *Dialectology, phonology, diachrony: Liverpool English realisations of PRICE and MOUTH*. Edinburgh: University of Edinburgh PhD Thesis.
- Clark, Lynn, Helen MacGougan, Jen Hay & Liam Walsh. 2016. "Kia ora. This is my earthquake story". Multiple applications of a sociolinguistic corpus. *Ampersand* 3. 13–20.
- Dann, Holly. 2019. *Productions and Perceptions of BATH and TRAP Vowels in Cornish English*. Sheffield: University of Sheffield PhD Thesis.
- Fromont, Robert & Jennifer Hay. 2012. LaBB-CAT: An annotation store. In *Proceedings of the Australasian Language Technology Association Workshop*, 113–117. University of Otago.
- Hall-Lew, Lauren, Claire Cowie, Catherine Lai, Nina Markl, Stephen Joseph McNulty, Shan-Jan Sarah Liu, Clare Llewellyn, Beatrice Alex, Zuzana Elliott & Anita Klingler. 2022. The Lothian Diary Project: sociolinguistic methods during the COVID-19 lockdown. *Linguistics Vanguard* 8(s3). 321–330.
- Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1(1). 97–120.
- Labov, William, Sharon Ash & Charles Boberg. 2008. *The atlas of North American English: Phonetics, phonology and sound change*. Berlin: Walter de Gruyter.
- Leemann, Adrian. 2021. Apps for capturing language variation and change in German-speaking Europe: Opportunities, challenges, findings, and future directions. *Linguistics Vanguard* 7(s1).
- Leemann, Adrian, Péter Jeszenszky, Carina Steiner, Melanie Studerus & Jan Messerli. 2020. Linguistic fieldwork in a pandemic: Supervised data collection combining smartphone recordings and videoconferencing. *Linguistics Vanguard* 6(s3). 1–16.
- Niedzielski, Nancy A. & Dennis R. Preston. 2003. *Folk Linguistics*. Berlin: Mouton de Gruyter.

- Ryan, Sadie Durkacz, Holly Dann & Rob Drummond. 2022. “Really this girl ought to be going to something better”: Rhoticity and social meaning in oral history data. *Language in Society*. 1–25.
- Rymes, Betsy & Andrea R Leone. 2014. Citizen Sociolinguistics: A New Media Methodology for Understanding Language and Social Life. *Working Papers in Educational Linguistics* 29(2). 25–43.
- Sneller, Betsy, Suzanne Evans Wagner & Yongqing Ye. 2022. MI Diaries: ethical and practical challenges. *Linguistics Vanguard* 8(s3). 307–319.
- Stuart-Smith, Jane. 1999. Glasgow: Accent and voice quality. In Paul Foulkes & Gerard J. Docherty (eds.), *Urban voices: Accent studies in the British Isles*, 203–222. London: Arnold.
- Trudgill, Peter. 1990. *The Dialects of England*. Oxford: Blackwell.
- Walsh, Liam, Jen Hay, Derek Bent, Jeanette King, Paul Millar, Viktoria Papp & Kevin Watson. 2013. The UC QuakeBox Project: Creation of a community-focused research archive. *New Zealand English Journal* 27. 20–32.
- Zellou, Georgia, Michelle Cohn & Bruno Ferenc Segedin. 2021. Age- and Gender-Related Differences in Speech Alignment Toward Humans and Voice-AI. *Frontiers in Communication* 5. 1–11.

Appendix

The interview consisted of 14 questions and an optional structured elicitation task. First, there were two introductory questions. The second of these was important, as while this strand of the project was open to anybody who lives in Greater Manchester, it was useful to know if somebody had lived in the same location for most of their lives.

1. *Tell me something about yourself. For example, what’s your name and how old are you?*
2. *Tell me a bit about where you grew up, and where you live now.*

Question 3 explored people’s attitudes towards the concept of “Greater Manchester”, to determine whether people from different demographics tended to identify more with the historic county that their borough would have been part of prior to 1974.

3. *How do you feel about the name ‘Greater Manchester’? Do you use it? Or do you prefer another name for your part of the country?*

Questions 4, 5 and 6 asked for people’s thoughts on their local area, the people in it, and how it compared to other areas of Greater Manchester.

4. *How do you feel about your local area within Greater Manchester? Is it a good place to live? Is there anything unique about it?*
5. *How would you describe the people that live in your community?*
6. *What do you think of other parts of Greater Manchester? Which areas and people are similar to yours, and which are different?*

Questions 7, 8 and 9 explicitly asked about participants’ own speech and how it relates to both the region, and to other people. Question 9 attempted to gather linguistic examples.

7. How would you describe the way you speak - your accent and dialect?

8. Is your accent typical of where you are from? How does it compare to the accent of your family and friends?

9. Do you think people can tell where you are from in Greater Manchester by the way you speak? For example, are there any pronunciations, words or grammar you use that are typical of where you live?

Questions 10, 11, 12 and 13 tried to explore the relationship between accent and identity beyond that of region. Question 11 asked participants to consider situations in which they might style-shift, and question 12 invited examples of prejudice.

10. Do you think they can tell anything else about you by the way you speak?

11. Do you think you speak differently in different situations and with different people? If so, in what ways?

12. Do you think the way you speak has ever caused you any problems? Or perhaps it has worked in your favour?

13. Do you think the way you speak is linked to who you are? If so, in what ways?

The final question simply asked how people feel about the way they speak.

14. Do you like the way you speak? Would you ever want to change it?

Adult participants were given all 14 questions in the same order, and they could see how far along they were by looking at the progress bar at the bottom of the screen (Figure 4). For participants under the age of 16, question 3 was removed, as it was felt that children might not be aware of the concept of Greater Manchester. Questions 4 and 9 were also adjusted to remove the term "Greater Manchester", and Question 6 was removed on the basis that children were less likely to know a lot about the different areas.