Research Article

International Journal of Qualitative Methods
Volume 23: 1–15
© The Author(s) 2024
DOI: 10.1177/16094069241247473
journals.sagepub.com/home/ijq
Sage

From "Listen and Repeat" to "Listen and Revise": How to Transcribe Interviews Offline Quickly and for Free Using Voice Recognition Software

Fabio Battaglia[1]

Abstract

Transcribing interviews is one of the most time-consuming and alienating tasks in qualitative research. Some have tried to bypass the problem by hiring external transcribers, which however can be very expensive. Whether due to the time and energy or the financial investment that transcribing requires, researchers therefore often self-impose a limit on the number of interviews that they conduct or even refuse to conduct interviews altogether. To prevent this and help reduce transcription fatigue, some scholars have developed the so-called "listen and repeat" technique. This involves speaking into a microphone what one hears through their headset while a voice recognition software transcribes word by word what it hears. However, this technique has many limitations and still requires a considerable amount of time and effort to be put by the researcher. This article introduces an alternative transcription technique which helps overcome these problems thanks to recent advancements of Artificial Intelligence in the field of voice recognition. Although the drawbacks and unintended consequences of Artificial Intelligence are often highlighted, this article explores its use for interview transcription showing that it can improve drastically the work and life of qualitative researchers. More specifically, this article introduces a transcription technique which allows to generate transcripts fully offline (avoiding in so doing the security concerns that the rising number of cloud-based transcription platforms often raise), rapidly and at little to no cost which one only needs to revise whilst listening to interview recordings, which is why I call this the "listen and revise" technique.

Keywords
transcription, interviews, artificial intelligence, voice recognition, qualitative research

Introduction

Transcribing interviews has been described as a 'boring drudgery' (Johnson, 2011, p. 92), a 'painful' (Sarkar, 2021, p. 141) and 'notoriously time-consuming and often tedious task' (McMullin, 2021, p. 141) that can be a 'nightmare' (Bailey, 2023, p. 112) and that is therefore 'often contracted out' (Horrocks & King, 2010, p. 119) for people 'underestimate the amount of time for taped interviews to be transcribed' (Bogdan et al., 2016, p. 170). Transcribing 1 hour of conversation can take indeed up to 8 (Morris, 2015) or 10 hours (McCartan & Robson, 2016). It is so taxing that scholars like Hest (2022, pp. 91–92) reported being 'in constant pain', feeling moving their 'fingers in cement' and eventually being diagnosed with 'a myofascial pain issue'. Unsurprisingly, research handbooks often therefore recommend 'to be cautious about transcription', underlining 'the limits rather than the strengths of the process' (Point & Baruch, 2023, p. 2).

[1] Department of Social Policy, London School of Economics and Political Science, UK

Corresponding Author:
Fabio Battaglia, Department of Social Policy, London School of Economics and Political Science, 2nd Floor, Old Building, Houghton Street, London WC2A 2AE, UK.
Email: f.battaglia@lse.ac.uk

Some have tried to bypass the problem by contracting the task out (see Point & Baruch, 2023). This may be a solution for them, but not for the transcribers themselves who will still need to transcribe interviews manually, not to mention the additional costs that hiring a professional transcriber requires. To prevent contracting transcription out and help reduce transcription fatigue, some scholars developed the so-called 'listen and repeat' technique (Park & Zeanah, 2005, p. 246). This involves listening to interview recordings and repeating each word in front of a microphone so that a voice recognition software such as Dragon (e.g. Estable, MacLean and Meyer, 2004; Hest, 2022; Matheson, 2007; Park & Zeanah, 2005) or MacSpeech Dictate (e.g. Fletcher & Shaw, 2011) can transcribe them. According to Matheson (2007, p. 557), this technique makes transcription 'less physically and mentally taxing' and for Dempster et al. (2015, p. 109) it is an 'impressive' way of transcribing. In terms of time savings, however, Park and Zeanah (2005) found virtually no benefits whereas Johnson (2011) found that transcribing this way took longer than typing manually. While this technique may thus make transcription easier, transcribing remains a time-consuming task, not to mention the time investment needed to train the software to recognise one's voice and the cost of the software itself (at the time of writing, October 2023, Dragon Professional costs $699[1]), the cost of 'implementing a foot pedal' (Tang, 2023, p. 299) to aid with transcription, voice fatigue and frustration (Park & Zeanah, 2005) and the fact that dictation needs to be conducted in a quiet place (Fletcher & Shaw, 2011; Hest, 2022; Matheson, 2007; Park & Zeanah, 2005), among other things.

This article introduces an alternative transcription technique which helps overcome these problems thanks to recent advancements of Artificial Intelligence (AI) in the field of voice recognition. The opportunities and challenges of using AI to help with different aspects of the research process are being debated increasingly (e.g. Alqahtani et al., 2023; Checco et al., 2021; Chubb et al., 2022). Although its drawbacks and unintended consequences are often highlighted, this article shows that its use for interview transcription can instead bring a multitude of benefits. More specifically, this article introduces a transcription technique which, using state-of-the-art voice recognition programmes that rely on Whisper's language models, allows to generate transcripts fully offline, rapidly and at little to no cost which one only needs to revise whilst listening to interview recordings, which is why I call this the "listen and revise" technique. Since no Internet connection is required, this technique helps to avoid the security concerns that the rising number of cloud-based transcription platforms often instead raise (Da Silva, 2021). Moreover, it helps to maximise efficiency and save an incredible amount of time and money, as well as to improve the research experience and potentially to increase the use of face-to-face interviews which scholars often limit or refuse to conduct altogether not only because of the time commitment (e.g. Erickson et al., 2004; Walker, 2006) and financial investment that transcribing requires, but also because of the physical and mental repercussions associated with it. To be sure, the technique in question is not flawless. Performance, for instance, varies from language to language and further improvements in that respect are needed. However, as this article shows the level of accuracy that voice recognition programmes can already achieve can change drastically the work and life of qualitative researchers, and therefore help make one of the most detested tasks in qualitative research only a memory of the past.

From "Listen and Repeat" to "Listen and Revise"

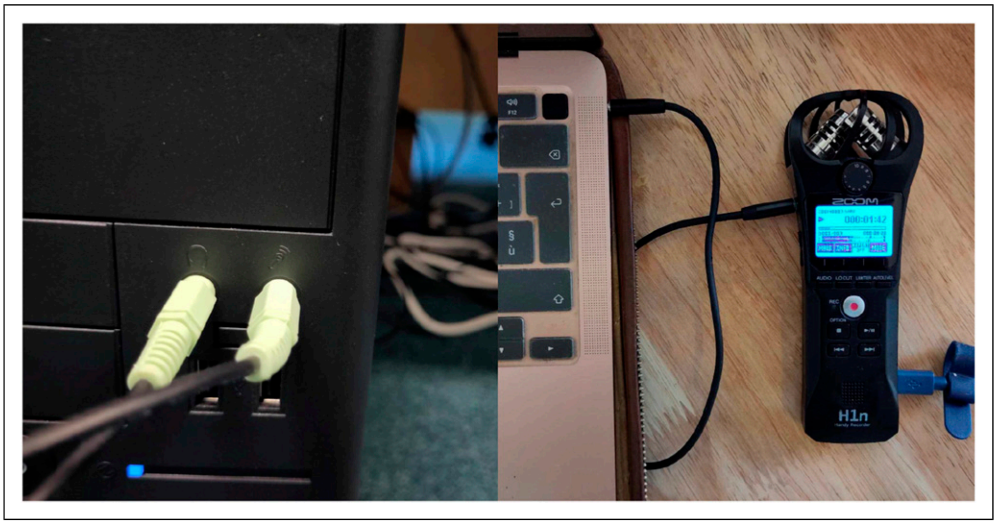
The "listen and repeat" technique (Park & Zeanah, 2005; see also Matheson, 2007) owes its name to the fact that, instead of transcribing interviews manually word by word, researchers can speak into a microphone what they hear through their headset as they play their recordings and let a voice recognition software transcribe for them, hence the alternative 'Parrot Method' (De Felice & Janesick, 2015, p. 1582). On the one hand, this helps reduce transcription times (Fletcher & Shaw, 2011), although less than one would expect (Park & Zeanah, 2005; see also Johnson, 2011). On the other hand, researchers using the "listen and repeat" technique have to transcribe interviews in a quiet – preferably empty – place (e.g. Fletcher & Shaw, 2011; Matheson, 2007; Park & Zeanah, 2005) for it is not just the interview that needs to be conducted in an empty venue, *but dictation, too*, in order to obtain good transcripts. Consequently, researchers transcribing interviews this way often have to isolate themselves (e.g. Hest, 2022; Park & Zeanah, 2005). Furthermore, the "listen and repeat" technique can lead to voice fatigue and frustration (Park & Zeanah, 2005) and it requires one to spend time to train the software to recognise their voice in the first place (e.g. Matheson, 2007).

Fortunately, there are three ways to circumvent these problems and transcribe interviews without having to listen to all recordings and repeat every word in front of a microphone. The first alternative simply entails connecting a device from which the recordings can be played to some external speakers, placing the latter as close to one's laptop's or phone's microphone as possible, play the recordings and let a voice recognition software transcribe. This, for instance, is similar to the strategy used by Da Silva (2021). However, just like in the "listen and repeat" technique, this means that other people, such as neighbours or colleagues, might be able to listen to the content of the interview – in fact, in this case they might be able to listen to the *actual* interview and hear participants' own voices – which is why to avoid privacy concerns one would need to transcribe in an isolated place. Whilst Da Silva (*ivi*, p. 5) lives in 'an attic room with a well-insulated door', it will not always be easy for researchers to find a similarly suitable place to transcribe. Additionally, transcripts will not necessarily be better than those generated with the traditional "listen and repeat" technique since the laptop's or phone's microphone may misinterpret or not capture all words being spoken (*ivi*). One could use Hi-Fi speakers to improve the sound quality, but this will not necessarily help generate better transcripts (*ivi*), to say nothing about the cost.[2]

The second and third alternatives have instead transformational results and are the subject of this article. I excogitated the first of such alternatives in 2017 and further refined it over the years to help me reduce transcription times. In a nutshell, this involves making one's device capture the audio coming from within itself as opposed to the one coming from external speakers or one's own voice. There are three ways through which this can be achieved. The first one is to use an old laptop or desktop computer. Modern devices tend to be very thin, with limited ports. Older laptops or desktop computers come instead with many ports and, more specifically, they usually have a headphone *and* a microphone port. By connecting the two with a male-to-male aux cable[3] (see the left side of Picture 1 below), one can make the internal microphone of their device capture the audio coming from within itself. By then using a voice recognition software that comes with live dictation and playing the interview recordings, one can get full transcripts without much effort, if any effort at all. The second way is to use a three-ring male-to-male aux cable[4] and connect a more modern laptop or desktop computer (which usually come with a combined audio input/output port rather than two individual ports) to any external device from which recordings can be played, such as the recorder itself (see the right side of Picture 1). The third way is to install instead a virtual audio device such as VB-CABLE[5] and configure it as the system's audio input and output (see Appendix A to learn how to do that). Installing such device is usually free and it essentially replaces the need of having a male-to-male aux cable, enabling one to transcribe interviews on tablets, too, which may not always have a headphone and/or microphone port. Moreover, by installing virtual audio devices like VB-CABLE on multiple devices (or by similarly using multiple aux cables to connect multiple devices), one can transcribe multiple interviews simultaneously therefore reducing even more drastically the time needed to complete the task.

Effectively, transcribing interviews this way does not require listening and repeating but rather reading the generated transcripts while listening to interview recordings and revising them when necessary, which is why I call this the "listen and revise" technique. Transcripts, indeed, are not always perfect and punctuation often needs to be added manually. Most importantly, if the original tape has background noises or interviewees did not speak clearly enough a lot of fine-tuning is usually needed. For this reason, it is critical to purchase a professional recorder – in my case, I regularly use a Zoom H1n with two stereo 90° microphones. However, this is not enough as what also contributes to obtaining good recordings (and consequently good transcripts) is placing the recorder as close to the interviewee as possible, as well as the emptiness of the place the interview is being recorded in (the emptier, the better). Unfortunately, in some cases one can only arrange interviews in public or busy areas. When this is the case, one should again make sure to place the recorder as close to the person as possible and programmes like Audacity[6] can be used afterwards to remove background noises.

There is, however, a further way to transcribe interviews following the "listen and revise" technique which can help obtain even more impressive results. Between September 2022 and October 2023, there have been significant advancements in the field of voice recognition which have made the use of a male-to-male aux cable or of a virtual audio device no longer necessary (albeit as I argue below, there are cases in which one may still want to use them). The launch of Whisper



Picture 1. Using a male-to-male aux cable to connect an old desktop computer's microphone port to its headphone port, on the left; using a three-ring male-to-male aux cable to connect a recorder to a more modern laptop, on the right (author's own pictures).

by OpenAI, in particular, has been a game changer. Therefore, a third and final alternative – and arguably the best way to transcribe interviews amongst all the options discussed in this article for reasons that will be dealt with more in depth in the next section – is to use Whisper's language models or programmes that rely on these models. Trained on 680,000 hours of audio data (Radford et al., 2022), Whisper is an open-source speech recognition system that supports transcription in 99 languages. It offers five different language packs (Tiny, Base, Small, Medium and Large), each with different turnaround times (the larger the package, the more accurate the transcript yet the longer it takes to generate it) and technical requirements (the larger the package, the higher the Random-Access Memory required to generate it). However, installing Whisper requires running some code and technical expertise, and using it, too, is not easy (see e.g. Stratvert, 2023). Therefore, being Whisper open source, this article focuses instead on programmes that have recently been developed that use Whisper's models but make them more accessible, SpeechPulse[7] (for Windows) and MacWhisper[8] (for Mac).[9]

Both programmes support live dictation and transcription of audio files in the same languages supported by Whisper and can add automatic punctuation. As an example, I wrote this very sentence using MacWhisper which also capitalised words for me. At the time of writing (October 2023), they come with free Tiny, Base (SpeechPulse) and Medium (MacWhisper) language models but offer the possibility, through a reasonably priced one-time payment of $19.95 and €29 (plus taxes),[10] respectively, to download the remaining models and transcribe multiple files at the same time (reducing even more the time needed to generate transcripts[11]), as well as to translate interviews, although this feature is currently underdeveloped.[12] Yet their greatest advantage, for reasons discussed in more detail in the next section, is that they process audio files and generate transcripts offline unlike Whisper which in principle requires an Internet connection to work.[13] Therefore, both programmes enable one to transcribe interviews quickly and for free (or at a very little cost), in many languages, at any time and from any part of the world.

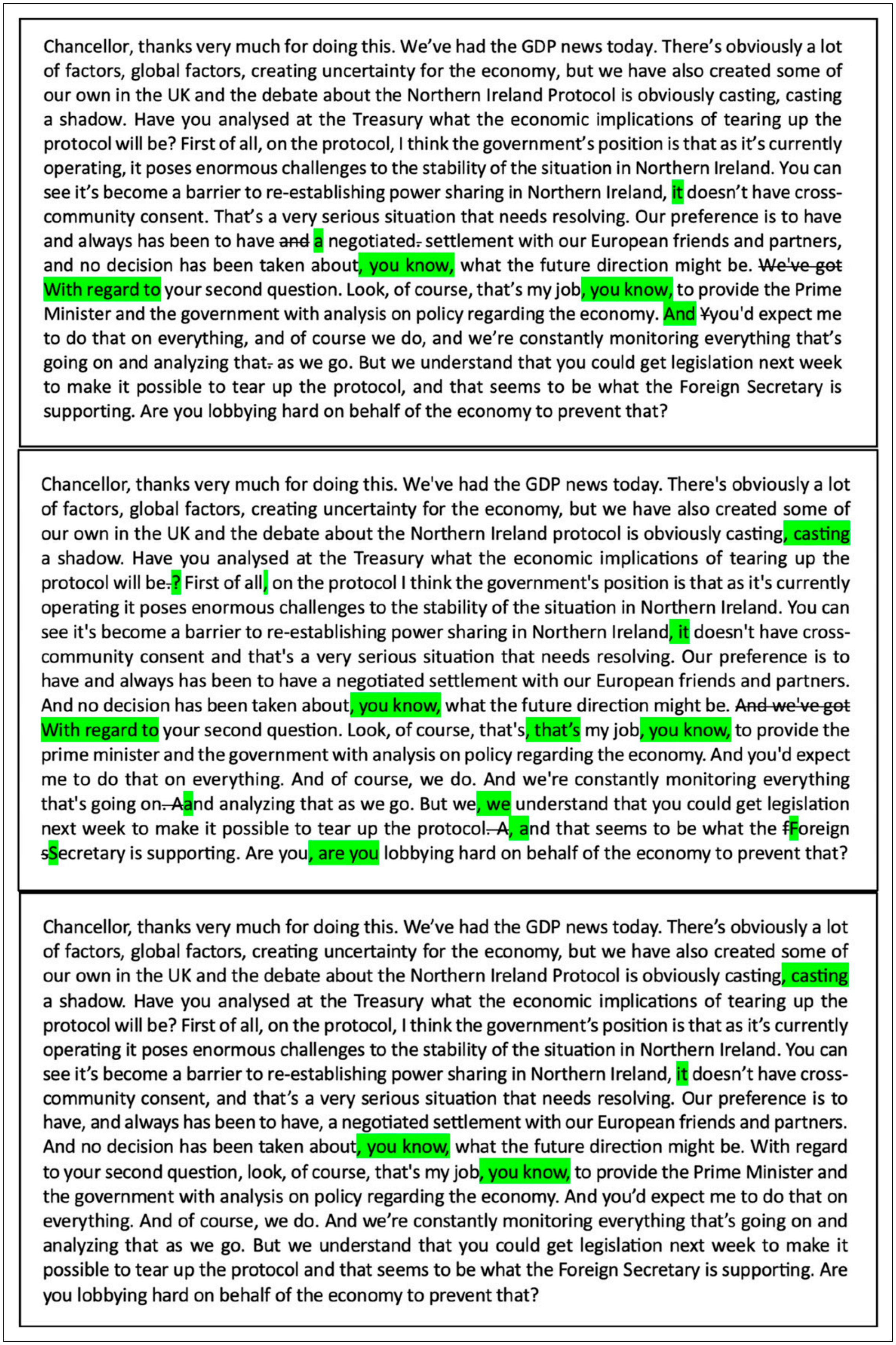
Below are two sample transcripts that I generated without Internet connection using SpeechPulse and MacWhisper of the first 90 seconds of a random interview that I found on YouTube of then Chancellor of the Exchequer of the United Kingdom, Rishi Sunak (Bloomberg, 2022). Indeed, it is worth noting that the "listen and revise" technique can not only be used to transcribe one's own interviews, but also *any* type of interview or spoken conversation more generally that can be played on one's device, as well as for live transcription. This makes this technique appetible not only to researchers, but also anyone interested in converting speech into text such as journalists, historians, lawyers or those who may struggle with typing. As one can see, the transcripts that both programmes generated even in Tiny and Base mode are impressively accurate and only required very minor corrections, highlighted in green, with those generated in Large mode requiring even fewer corrections (Figure 1). As for the time needed to generate the transcripts, transcribing via live dictation using VB-CABLE took about 95 seconds (this being the length of the interview excerpt in question plus a few seconds needed to generate the transcript). It took instead only ca. 42 seconds for both programmes to transcribe *the whole* 17-minute interview when uploading the file directly. Moreover, they managed to transcribe the voices of both the interviewer and the interviewee, despite them talking right after the other. Therefore, provided each person speaks clearly and that the audio is of good quality, the number of interviewees does not seem to be a problem (in contrast with the "listen and repeat" technique which requires one to train the software to recognise their own voice). Gone are thus the days when 'providing accurate transcriptions of long blocks of actual human conversation' was deemed 'beyond the abilities of even today's most advance software' (Jarnow, 2016, n. p.).

Benefits and Limitations of Using Voice Recognition Programmes for Interview Transcription

The "listen and revise" technique saved me an incredible amount of time. It is estimated again that 1 hour of conversation takes up to 8 (Morris, 2015) or 10 hours (McCartan & Robson, 2016) to transcribe, with some scholars reporting even longer times.[14] The average interview in my research lasts 1 hour; considering that I spend on average 1.5 hours "transcribing" per hour of material, this means, assuming that it takes 9 hours to transcribe every hour of conversation, that I have been able to cut transcription times by 83%. This is an incredible result, especially since as we have seen transcribing is often considered one of the most daunting tasks in qualitative research and arguably one of the reasons why scholars and students alike often limit the number of interviewees in their samples or decide not to conduct face-to-face interviews altogether (e.g. Erickson et al., 2004; Walker, 2006).

The drastic reduction in the time needed to generate transcripts is not the only benefit of the "listen and revise" technique. Indeed, this technique does not only allow to transcribe multiple interviews at the same time effortlessly – it also enables one to do other things whilst interviews are being transcribed. This can be devoting more time to analysing data, sending e-mails to arrange even more interviews or, perhaps most importantly, undertaking paid work and simultaneously avoid contracting transcription out. For instance, instead of spending time transcribing I have been able to work to raise money for my research and save at least £7,200 on professional transcribing services.[15] Additionally, one can avoid requesting financial resources to hire external transcribers when applying for funding and that money can be used to fund other projects instead.

Yet the greatest benefit for me is the positive impact that all this has had on my mental health and the fact that I have been able to avoid all the negative aspects usually associated with



Chancellor, thanks very much for doing this. We've had the GDP news today. There's obviously a lot of factors, global factors, creating uncertainty for the economy, but we have also created some of our own in the UK and the debate about the Northern Ireland Protocol is obviously casting, casting a shadow. Have you analysed at the Treasury what the economic implications of tearing up the protocol will be? First of all, on the protocol, I think the government's position is that as it's currently operating, it poses enormous challenges to the stability of the situation in Northern Ireland. You can see it's become a barrier to re-establishing power sharing in Northern Ireland, it doesn't have cross-community consent. That's a very serious situation that needs resolving. Our preference is to have and always has been to have ~~and~~ a negotiated~~.~~ settlement with our European friends and partners, and no decision has been taken about, you know, what the future direction might be. ~~We've got~~ With regard to your second question. Look, of course, that's my job, you know, to provide the Prime Minister and the government with analysis on policy regarding the economy. And ~~Y~~you'd expect me to do that on everything, and of course we do, and we're constantly monitoring everything that's going on and analyzing that~~.~~ as we go. But we understand that you could get legislation next week to make it possible to tear up the protocol, and that seems to be what the Foreign Secretary is supporting. Are you lobbying hard on behalf of the economy to prevent that?

Chancellor, thanks very much for doing this. We've had the GDP news today. There's obviously a lot of factors, global factors, creating uncertainty for the economy, but we have also created some of our own in the UK and the debate about the Northern Ireland protocol is obviously casting, casting a shadow. Have you analysed at the Treasury what the economic implications of tearing up the protocol will be~~.~~? First of all, on the protocol I think the government's position is that as it's currently operating it poses enormous challenges to the stability of the situation in Northern Ireland. You can see it's become a barrier to re-establishing power sharing in Northern Ireland, it doesn't have cross-community consent and that's a very serious situation that needs resolving. Our preference is to have and always has been to have a negotiated settlement with our European friends and partners. And no decision has been taken about, you know, what the future direction might be. ~~And we've got~~ With regard to your second question. Look, of course, that's, that's my job, you know, to provide the prime minister and the government with analysis on policy regarding the economy. And you'd expect me to do that on everything. And of course, we do. And we're constantly monitoring everything that's going on~~. A~~and analyzing that as we go. But we, we understand that you could get legislation next week to make it possible to tear up the protocol~~. A~~, and that seems to be what the ~~f~~Foreign ~~s~~Secretary is supporting. Are you, are you lobbying hard on behalf of the economy to prevent that?

Chancellor, thanks very much for doing this. We've had the GDP news today. There's obviously a lot of factors, global factors, creating uncertainty for the economy, but we have also created some of our own in the UK and the debate about the Northern Ireland Protocol is obviously casting, casting a shadow. Have you analysed at the Treasury what the economic implications of tearing up the protocol will be? First of all, on the protocol, I think the government's position is that as it's currently operating it poses enormous challenges to the stability of the situation in Northern Ireland. You can see it's become a barrier to re-establishing power sharing in Northern Ireland, it doesn't have cross-community consent, and that's a very serious situation that needs resolving. Our preference is to have, and always has been to have, a negotiated settlement with our European friends and partners. And no decision has been taken about, you know, what the future direction might be. With regard to your second question, look, of course, that's my job, you know, to provide the Prime Minister and the government with analysis on policy regarding the economy. And you'd expect me to do that on everything. And of course, we do. And we're constantly monitoring everything that's going on and analyzing that as we go. But we understand that you could get legislation next week to make it possible to tear up the protocol and that seems to be what the Foreign Secretary is supporting. Are you lobbying hard on behalf of the economy to prevent that?

Figure 1. The transcripts generated offline by SpeechPulse (top) and MacWhisper (middle) in Tiny and Base Mode, respectively, and by MacWhisper (bottom) in Large mode of the first 90 seconds of Bloomberg (2022), after manual revision (author's own screenshots).

both manual transcription and the “listen and repeat” technique, which can be ‘equally dull’ (Johnson, 2011, p. 95) and arguably more challenging than manual transcription. Indeed, as discussed above researchers using the “listen and repeat” technique have to transcribe interviews in a quiet place (e.g. Fletcher & Shaw, 2011; Matheson, 2007; Park & Zeanah, 2005) and often have to isolate themselves (e.g. Hest, 2022; Park & Zeanah, 2005). Since the technique I outlined above does not require to speak into a microphone, one does not need to be alone when transcribing. In fact, I often transcribed interviews while working with my colleagues or friends, therefore avoiding the self-imposed isolation that the “listen and repeat” technique requires.

Using voice recognition programmes to transcribe interviews is not, however, flawless. To begin with, the final amount of time that one needs to spend revising transcripts depends ultimately on what one’s research aims are. In principle, one can simply add any missing punctuation and correct mistakes when needed. Someone doing e.g. discourse analysis, instead, will need to spend more time fine-tuning, but the transcription technique outlined above will still save them a significant amount of time, not to mention that as we have seen voice recognition programmes can also add automatic punctuation, thus further reducing the time needed to obtain good verbatim transcripts. Relatedly, “ums” and “ers” may not be captured but can be added in the revision phase if needed. It is also worth noting that, regarding Whisper’s language models specifically, performance varies from language to language (see GitHub.com, 2023b). As it will be discussed below, this is why in some cases one may want to use other programmes, as these may be able to generate better transcripts.

Another point worth highlighting is that transcripts may be less accurate if people are speaking over each other. This, however, does not mean that voice recognition programmes cannot help transcribing focus groups or situations in which more than one informant is involved. When the “listen and repeat” technique was first developed, voice recognition programmes could only recognise one voice at a time (e.g. Matheson, 2007). As we saw above, however, they are now able to recognise multiple voices and may therefore help transcribe conversations involving more than one informant provided that one person speaks at a time, close to the recorder and clearly enough. After all, speech recognition software is being used in (crowded) parliaments like the Japanese Diet, where since 2011 an internal software is helping transcribe committee and plenary meetings with an overall accuracy of 90% (Kawahara, 2012; see also Kawahara et al., 2021).

There are some potential unintended consequences of using voice recognition programmes that should be taken into consideration. According to Bolden (2015, p. 277), these programmes turn researchers into ‘simple observers of interactions’ since they do not engage in a ‘close and careful listening’ of their recordings. It is important to note that Bolden was writing specifically for those doing conversational analysis, hence their concerns will be less relevant to those not taking an analytic approach to transcription. At any rate, such concerns are built on the premise that transcription ends as soon as transcripts are generated. This, however, is far from accurate which is exactly why I have called it the “*listen* and *revise*” technique: this way of transcribing does *not* eliminate the need for listening carefully to interview recordings – it simply removes the need for producing good initial transcripts whose accuracy one has now actually *more* time to review. Second, Bolden (*ibidem*) suggested that researchers may end up privileging data that work best with these programmes, conducting interviews with little interactions and ‘where all participants speak a standard dialect of a language’. As shown and argued above, voice recognition programmes can recognise multiple voices and in quite interactive contexts such as televised interviews or parliamentary debates. Moreover, the number of languages covered is already vast and increasing, and several language variants are also available. In general, rather than having ‘a negative impact’ on the study of ‘lesser-known languages’ (*ibidem*), it seems to me that voice recognition programmes can instead stimulate and help researchers transcribe interviews in languages that they may not be fluent in and help *preserve* lesser-known languages by making them available to future generations of researchers and transcribers.

Finally, the greatest problem with using voice recognition programmes for interview transcription comes from the potential collection, storage of, access to, use and retention of data (that is, one’s recordings and any related transcripts) by the companies that administer such programmes, as well as any third parties (Da Silva, 2021). This is because even when recordings do not need to be uploaded and live transcription is available, companies may still receive and store voice inputs and any associated transcripts, sometimes in countries that differ from the one where the researcher is based in, and these may be reviewed by humans and used for training or other purposes. Such concerns have been raised not only with regard to academic research but also journalistic work (Shelton & Grauer, 2022), with journalists being asked questions about sensitive information contained in their transcripts by the companies owning transcription platforms (Kine, 2022). To understand whether Da Silva’s (2021) concerns apply to our case or not it is necessary to distinguish between two categories of voice recognition programmes: those that generate transcripts offline, and cloud-based platforms which require instead data to leave one’s device for transcription to be performed and therefore also an Internet connection to do that. SpeechPulse (2023b, n. p.) ‘works fully offline’. Similarly, in the case of MacWhisper (n. d. a, n. p.) ‘[a]ll transcription is done on your device, no data leaves your machine’, in line with the product’s remarkably clear Privacy Policy (‘[w]e don’t want to know anything about you’ – MacWhisper, 2023, n. p.). Both programmes thus belong to the former category – indeed, the purpose of this article was not to find a way to mitigate potential security concerns, but rather to develop a way of transcribing that would *avoid such concerns in the first place*. Furthermore, since both programmes do not require an Internet connection to transcribe, their use also helps avoid the 'risk of data being intercepted', which instead can happen 'during upload to cloud-based services' (Da Silva, 2021, p. 4).

Having said that, one can never be safe enough and given that interviews will contain personal or sensitive information, this article's recommendation is to *always* transcribe offline on an ad hoc device where, after downloading all language packs needed and entering any activation key to unlock additional features, Internet has been permanently disabled, i.e. a standalone device that is used only for transcribing and other offline tasks.[16,17] This is where one of SpeechPulse's and MacWhisper's greatest comparative advantage lies: since they are not subscription-based, once they are installed they no longer require an Internet connection to renew their license, therefore offering a lifetime solution for offline transcription that does not also require any future costs. Since the device should only be used offline, there is of course the question of how any programme updates could be installed, given that new languages may be added in the future or that performance of current languages may be improved. However, this can easily be solved by downloading any new future language packs or programme versions – or even completely new programmes that currently do not exist – and transferring them to the offline device via an USB drive or SD card. Finally, an even further level of security should be added by encrypting all relevant files and installing a firewall software such as GlassWire or Portmaster (for Windows) or Little Snitch (for Mac) or any equivalent programme before airgapping the device and using it to block any potential connection attempts from the programme that one is using.[18]

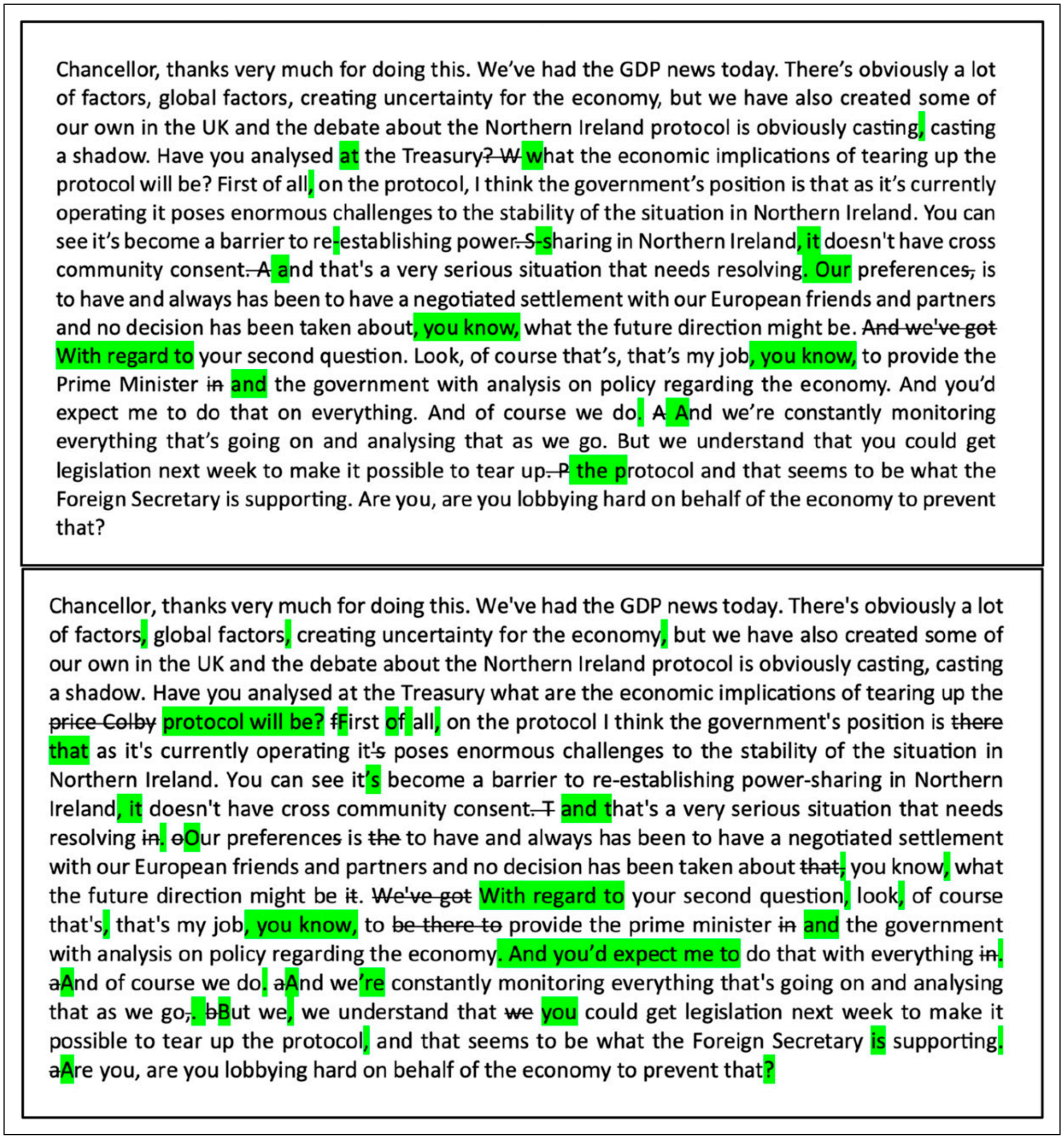
It is worth noting that to run Whisper's language models – and, consequently, SpeechPulse and MacWhisper – a device with certain technical requirements is needed.[19] It may not always be possible to either buy or keep a new laptop only for transcription, but purchasing a second-hand one might do the trick.[20] Rather than a laptop, one could purchase a tablet instead that meets the minimum system requirements of either programme, therefore reducing costs even further.[21] Moreover, since such a device could be used endlessly to transcribe any future interviews or video/audio files, its cost can be spread over time. Additionally, one does not need to use Internet data since transcription is performed on the device and the device itself could be re-sold after use, although one should ensure to have permanently deleted all data before passing this over to someone else.[22]

All in all, using programmes that allow offline transcription and rely on Whisper's language models such as SpeechPulse and MacWhisper is by far at the moment the best way to transcribe interviews. Such programmes enable one to transcribe at any time and from any part of the world in various languages for free (or at a very little cost through a one-off payment) and above all to process data on the device. If not in possession already, buying an ad hoc device that is disconnected from the Internet and used only for transcription and other offline tasks will require a certain investment at the beginning. However, there are ways to minimise costs, both in the present (e.g. purchasing a tablet or second-hand laptop, or a second-hand tablet) and in the future (e.g. re-selling the device when no longer needed).

Potential Alternative Ways to Transcribe Interviews

Instead of using SpeechPulse or MacWhisper, an alternative option could be to use a word editing programme, as most of these programmes have added dictation to their suite of features in recent years. As noted above, the reason why one may want to use such programmes is because these may be able to generate better transcripts, for instance when transcribing to a language for which Whisper's language models' performance is lower. These programmes may also be used when one is not transcribing interviews but e.g. publicly available videos or audio files which may not require to take precautionary measures like the ones outlined above to ensure that transcription is fully performed offline. Windows users can use Word's Dictate option which supports ca. 39 languages currently plus several language variants. Word supports both live dictation and the direct upload of files, it can add automatic punctuation and its use is free. To be sure, one needs to be subscribed to Microsoft 365 for the feature to be available in the first place, but researchers often have free access to Microsoft 365 through their employer. Even if they do not, a basic personal subscription to Microsoft 365 currently costs £59.99 in the United Kingdom when billed annually,[23] hence it would still be cheaper than subscribing to other platforms available on the market (see below). If in possession of an Apple device, instead, one could use Apple's Pages, which also has a Dictation option available. Pages, too, can add automatic punctuation and it can transcribe in different languages (ca. 32 currently plus language variants). Contrarily to Word, however, it only supports live dictation (meaning that the only way to transcribe following the "listen and revise" technique outlined in this article is by using either a male-to-male aux cable or a virtual audio device as mentioned above – see Appendix A and B for instructions and notes) and, above all, it does not require any Internet connection to work. Below are sample transcripts that I generated using Word (online) and Pages (offline) in combination with VB-CABLE of the first 90 seconds of the Sunak interview. As it can be seen (Figure 2), the transcripts generated by both are very accurate, albeit slightly less so than the ones generated by SpeechPulse and MacWhisper.

Since Word's Dictate feature requires an Internet connection to function – i.e. data are sent to Microsoft – it is important to be mindful of the security concerns that Da Silva (2021) raised. Likewise, whilst Pages can process audio files and generate transcripts offline, this is only the case if the user has a specific device and if they have opted out of sharing their



Chancellor, thanks very much for doing this. We've had the GDP news today. There's obviously a lot of factors, global factors, creating uncertainty for the economy, but we have also created some of our own in the UK and the debate about the Northern Ireland protocol is obviously casting, casting a shadow. Have you analysed at the Treasury~~? W~~ what the economic implications of tearing up the protocol will be? First of all, on the protocol, I think the government's position is that as it's currently operating it poses enormous challenges to the stability of the situation in Northern Ireland. You can see it's become a barrier to re-establishing power~~. S~~-sharing in Northern Ireland, it doesn't have cross community consent~~. A~~ and that's a very serious situation that needs resolving. Our preferences~~,~~ is to have and always has been to have a negotiated settlement with our European friends and partners and no decision has been taken about, you know, what the future direction might be. ~~And we've got~~ With regard to your second question. Look, of course that's, that's my job, you know, to provide the Prime Minister ~~in~~ and the government with analysis on policy regarding the economy. And you'd expect me to do that on everything. And of course we do. ~~A~~ And we're constantly monitoring everything that's going on and analysing that as we go. But we understand that you could get legislation next week to make it possible to tear up~~. P~~ the protocol and that seems to be what the Foreign Secretary is supporting. Are you, are you lobbying hard on behalf of the economy to prevent that?

Chancellor, thanks very much for doing this. We've had the GDP news today. There's obviously a lot of factors, global factors, creating uncertainty for the economy, but we have also created some of our own in the UK and the debate about the Northern Ireland protocol is obviously casting, casting a shadow. Have you analysed at the Treasury what are the economic implications of tearing up the ~~price Colby~~ protocol will be? ~~f~~First of all, on the protocol I think the government's position is ~~there~~ that as it's currently operating it~~'s~~ poses enormous challenges to the stability of the situation in Northern Ireland. You can see it's become a barrier to re-establishing power-sharing in Northern Ireland, it doesn't have cross community consent~~. T~~ and that's a very serious situation that needs resolving ~~in~~. ~~o~~Our preference~~s~~ is ~~the~~ to have and always has been to have a negotiated settlement with our European friends and partners and no decision has been taken about ~~that,~~ you know, what the future direction might be ~~it~~. ~~We've got~~ With regard to your second question, look, of course that's, that's my job, you know, to ~~be there to~~ provide the prime minister ~~in~~ and the government with analysis on policy regarding the economy. And you'd expect me to do that with everything ~~in~~. ~~a~~And of course we do. ~~a~~And we're constantly monitoring everything that's going on and analysing that as we go~~,~~. ~~b~~But we, we understand that ~~we~~ you could get legislation next week to make it possible to tear up the protocol, and that seems to be what the Foreign Secretary is supporting. ~~a~~Are you, are you lobbying hard on behalf of the economy to prevent that?

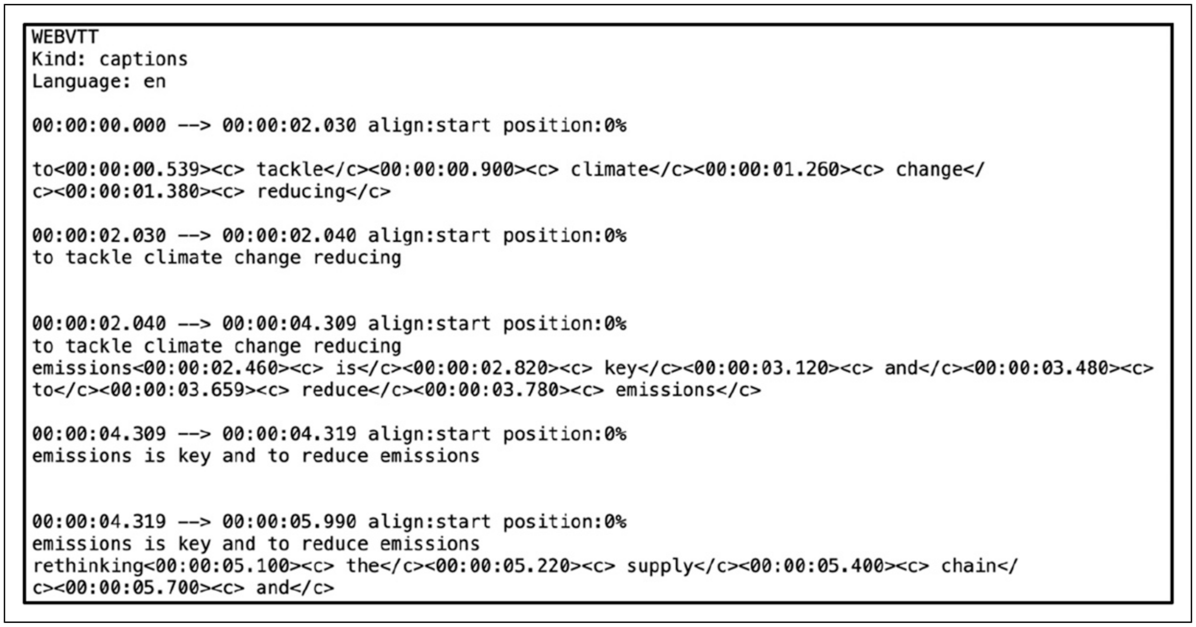
Figure 2. The transcripts generated by Word (top) and Pages (bottom) of the first 90 seconds of Bloomberg (2022), after manual revision (author's own screenshots).

speech utterances and transcripts with Apple (see Appendix A and B). Evaluating the potential use of Word and Pages in light of their privacy policies goes beyond the scope of this article, due to space constraints but also and above all to the fact that, given that interviews will contain personal or sensitive information, the aim here was again that of finding ways to transcribe *fully offline*. However, as noted above there may be cases in which researchers may want to use Word or Pages. Although some guidance on the use of Pages is provided in Appendix B, Word was not covered as it currently requires an Internet connection to transcribe. Researchers wishing to use Word are therefore encouraged to consult a cybersecurity specialist to ensure compliance with any data protection regulations in force in their country.

It is worth acknowledging that there are also other programmes or platforms the use of which scholars have explored in the literature. An example of such platforms is YouTube (Bokhove & Downey, 2018). After creating a video with a blank image that stays on the screen for as long as the recording's duration, one can upload this to YouTube which, after processing it, generates automated captions which one can then download and review (*ivi*). This way of transcribing, however, is the least recommended amongst all the ones discussed in this article. First, it does not solve the privacy issues raised by Da Silva (2021) – in fact, it possibly complicates them as one needs to upload a video making this thus subject to Google's Privacy Policy. Second, it requires every time creating a video with a blank image and the very act of uploading it, thus extra unnecessary and time-consuming steps. Third and last, the transcript produced by YouTube includes timestamps and other irrelevant information (see Figure 3). As Bokhove and Downey (2018) noted, another programme can be used to remove them, but this would require even further unnecessary and time-consuming steps.

A further potential alternative entails the use of live conferencing programmes. These programmes, which include Zoom and Microsoft Teams, are being increasingly used to conduct interviews remotely (e.g. Richardson et al., 2021) and they, too, help make transcription easier as they allow users to download the transcripts of their calls. However, given that these platforms require an Internet connection to work, one needs to be mindful again of the privacy issues that their use might raise (Da Silva, 2021). Researchers are therefore encouraged to check the privacy policy of the products they wish to use, as well as to consult a cybersecurity specialist, before doing so. Nevertheless, irrespective of these platforms' privacy policies, transcripts generated from online interviews are less likely to be as accurate as those generated from in-person ones. This is because whilst during an in-person interview the researcher can affect the quality of the transcript by choosing an empty venue, bringing a good quality recorder and placing it close to the interviewee(s), the accuracy of transcripts in the case of online interviews depends instead on the quality of the audio of the computer's or headset's microphone, and of the Internet connection, of *all* the people involved, and especially of the interviewee(s) – which the researcher cannot in this case control – as well as on the emptiness of the venues in which both the interviewer and the interviewee(s) are. And – it goes without saying – the more people one interviews at the same time, the more these factors will influence the possibility of generating accurate transcripts. To be sure, there will be times in which doing interviews online is the only available option, be it for environmental, logistical or financial reasons, or due to health reasons as it was the case for many during the Coronavirus disease pandemic. However, considering the problems above, using these platforms for the mere purpose of generating transcripts is in my opinion not recommended if one can conduct an in-person interview instead.

Finally, as noted at the beginning other voice recognition programmes that scholars have used and discussed in the literature are Dragon (e.g. Estable, MacLean and Meyer, 2004; Hest, 2022; Matheson, 2007; Park & Zeanah, 2005) and MacSpeech Dictate (e.g. Fletcher & Shaw, 2011). Other platforms the use of which has been reviewed in the literature are also Otter.ai (Bourgeault & Corrente, 2022) and NVivo Transcription (di Gregorio, 2022). But these are not the only options. Over the years, the number of platforms offering automatic transcription have indeed mushroomed: examples include Amberscript, Descript, Happy Scribe, Notta, Simon Says, Speechnotes and Trint, to mention but a few. Since the vast majority of these programmes or platforms do not perform transcription on the device but require an Internet connection not only to function but in some cases to also renew their license, as well as the purchase of a subscription or pay-as-you-go plan in the first place (which can be expensive[24]), they were not the subject of this study exactly because the aim was again to identify a way to transcribe interviews that



```
WEBVTT
Kind: captions
Language: en

00:00:00.000 --> 00:00:02.030 align:start position:0%

to<00:00:00.539><c> tackle</c><00:00:00.900><c> climate</c><00:00:01.260><c> change</
c><00:00:01.380><c> reducing</c>

00:00:02.030 --> 00:00:02.040 align:start position:0%
to tackle climate change reducing


00:00:02.040 --> 00:00:04.309 align:start position:0%
to tackle climate change reducing
emissions<00:00:02.460><c> is</c><00:00:02.820><c> key</c><00:00:03.120><c> and</c><00:00:03.480><c>
to</c><00:00:03.659><c> reduce</c><00:00:03.780><c> emissions</c>

00:00:04.309 --> 00:00:04.319 align:start position:0%
emissions is key and to reduce emissions


00:00:04.319 --> 00:00:05.990 align:start position:0%
emissions is key and to reduce emissions
rethinking<00:00:05.100><c> the</c><00:00:05.220><c> supply</c><00:00:05.400><c> chain</
c><00:00:05.700><c> and</c>
```

Figure 3. An example of a YouTube-generated transcript (author's own screenshot).

would *never* require an Internet connection, except for initially downloading the programme or activating any licenses. If intending to use any of these cloud-based programmes or platforms, researchers are encouraged to consult their privacy policies to check how their data are going to be managed and whether and what type of data are going to be collected in the first place. However, a word of advice is needed here because such documents are often vague or written in formal, technical and complicated language, which makes them hard to understand, to say nothing of the fact that it is often hard to get clear responses from companies when they are approached for clarifications, if they reply at all.[25] Privacy policies are also updated regularly and will therefore change over time, thus requiring constant monitoring and the capacity to understand each time the impact that any changes will have on the way that data are collected, stored and used, among other things. Furthermore, it is worth noting that several other platforms, especially those that offer free voice-to-text transcription such as SpeechTexter or LilySpeech, actually rely on Google to turn speech into text[26] and may perform 'market research in the background' (LilySpeech, 2020, n. p.) hence one needs to be careful, in the attempt to find a cheaper solution, not to end up finding themselves being subject to two different companies' privacy policies rather than just one. This is where a final, further strength of the "listen and revise" technique lies: although one should always consult the privacy policies of the products that they are using, if transcription is performed following the directions provided in this article (i.e. using a standalone device that has been permanently disconnected from the Internet, where files have been encrypted and any potential connection attempts blocked), transcripts are generated offline, hence not only will no data leave the device but these will also not be subjected to future privacy policy changes.

Conclusions

To date, research handbooks (e.g. Bailey, 2023; Brinkmann, 2022; Brinkmann & Kvale, 2018; Flick, 2018) still provide no alternatives to manual transcription or, when they do, they only mention the "listen and repeat" technique (e.g. Silver & Lewins, 2020). Unsurprisingly, researchers still therefore see transcribing as one of the most daunting and time-consuming tasks of qualitative research, and often try to bypass it by hiring external transcribers. Thanks to recent advancements of AI in the field of voice recognition, transcribing interviews is however no longer a painful task, let alone a time-consuming one. Although the drawbacks of using AI in research are often highlighted, this article showed that its use for interview transcription can instead improve drastically the work and life of qualitative researchers. More specifically, this article introduced a transcription technique which, using state-of-the-art voice recognition programmes that rely on Whisper's language models, allows to generate transcripts fully offline, rapidly and at little to no cost which one only needs to revise whilst listening to interview recordings, which is why I have called this the "listen and revise" technique.

This technique helped me reduce transcription times by 83% and save at least £7,200 over the years, but there are also many other benefits. To begin with, all the time that is saved can be dedicated to conduct more interviews and generate more information, or to analyse data. Similarly, all the money saved can be used to fund extra fieldwork or to hire research assistants to help with other tasks. Possibly even better, one can avoid requesting financial resources to hire external transcribers when applying for funding and that money can be used to fund other projects instead. Being transcribing very alienating, repetitive and frustrating, this technique also allows to improve the researcher's physical and mental health – or at least it helps not to worsen it – and to make the research process a better experience overall. Researchers using the "listen and repeat" technique have to transcribe interviews in a quiet place, and therefore often isolate themselves (e.g. Hest, 2022; Park & Zeanah, 2005). The transcription technique outlined in this article helps prevent that, as one does not need to speak into a microphone and consequently to work in a room alone. Provided people spoke clearly enough and as close to the recorder as possible, one can also transcribe interviews in which more than one informant was involved and do so in many languages, thus facilitating transcription in cases in which researchers did not conduct interviews in their native language. This in turn can make them feel less worried about conducting such interviews and increase the linguistic diversity of qualitative research, as well as help preserve lesser-known languages by making them available to future generations of transcribers. Furthermore, programmes such as SpeechPulse and MacWhisper can add automatic punctuation and capitalise words, too, reducing the time needed to generate accurate transcripts even further. It is also worth noting that the "listen and revise" technique can be used to transcribe any type of spoken conversation that can be played on one's device and not just interviews. Those who can use it and benefit from it are thus not just university students and scholars, but also journalists, historians, lawyers and, broadly speaking, anyone interested in converting speech into text or those who may struggle with typing.

Yet possibly the greatest advantage is the fact that files are processed, and transcripts generated, fully offline on the device. This is possible thanks to recent advancements in the field of voice recognition, particularly the launch of Whisper by OpenAI and the subsequent development of programmes such as SpeechPulse and MacWhisper which make Whisper's language models more accessible and easier to use. When operated on an ad hoc device that is kept offline, both programmes can generate impressively accurate transcripts quickly and at little to no cost, offering an affordable lifetime solution to the tediousness of manual transcription. This ensures not only that interview recordings and transcripts are not leaving the device and helps prevent data interception as well as using Internet data, but it also avoids the need of constantly monitoring companies' and their products' (complicated) privacy policies, which is where further strengths of this technique lie. All in all, provided adequate measures like the ones outlined in this article are taken to address the privacy concerns that Da Silva (2021) duly raised, the multitude of benefits brought by voice recognition programmes clearly outweigh the limitations, which are at this point hard to even identify.

Artificial Intelligence should neither be praised nor condemned a priori, but its uses explored and evaluated. As far as the use of voice recognition programmes for interview transcription is concerned, we can firmly say that such programmes have started a Copernican revolution in qualitative research. All things considered, it is indeed no doubt that the "listen and revise" technique presented in this article is a breakthrough in the field, as it will not only make research more affordable, efficient and enjoyable, as well as less alienating, repetitive and isolating, but it will also prevent researchers from voluntarily limiting their interview samples or from deciding to not conduct interviews altogether. In turn, this will help gather more information that would have not been so otherwise, helping researchers strengthen or discover new findings and hopefully contribute to make qualitative research stronger in the eyes of those who regard it as an inferior method of enquiry.

Appendix A

Instructions for Apple Users

The following instructions apply to silicon devices running Ventura and Intel-based devices running Mojave and may therefore not be applicable if using devices with newer or older operating systems (see further notes in Appendix B).

1. Configuring VB-CABLE. After downloading and installing VB-CABLE, you need to configure this as the system's audio output. To do so, click on System Preferences > Sound > Output and then select VB-CABLE. You will also need to configure it as the system's audio input. To do so, click on System Preferences > Sound > Input and then select VB-CABLE. Remember to restore the default output and input sources after you are done with transcription or else you will not be able to hear any sound from your device.

2. Turning Improve Siri and Dictation off (for Apple Silicon Devices). To turn Improve Siri and Dictation off, simply click on System Preferences > Privacy & Security > Analytics & Improvements > and then tap to turn Improve Siri & Dictation off.

3. Turning Enhanced Dictation on (for Intel-based Devices). To turn Enhanced Dictation on, simply click on System Preferences > Keyboard > Dictation and then tick the Use Enhanced Dictation box.

4. Troubleshooting: If Live Dictation is Not Working. If, when using a male-to-male aux cable or a virtual audio device and transcribing through live dictation the latter is not working, make sure that the audio is not being played as mono. To do that, click on Accessibility > Audio > and make sure that the Play stereo audio as mono box is unticked.

5. Troubleshooting: If Dictation is Not Working (Only Applies if Using Pages). Make sure that VB-CABLE is selected as Dictation's microphone source (System Preferences > Keyboard > Dictation > Microphone source). If Dictation still does not work, add another language (System Preferences > Keyboard > Dictation > Language) and, after clicking on Edit > Start Dictation within Pages, change the language from the pop-up menu to the one that you have just added and then change it back again to the default one.

Please note that while Dictation on a silicone device transcribes uninterruptedly, when using Enhanced Dictation on an Intel-based device running High Sierra and Mojave I had to take a break and pause the audio every 15–20 seconds to let Pages process the audio and generate the transcript or else Enhanced Dictation would not work, hence using Enhanced Dictation on an device that has Intel processors may require the researcher to stay in front of the laptop for the duration of the interview.

Instructions for Windows Users

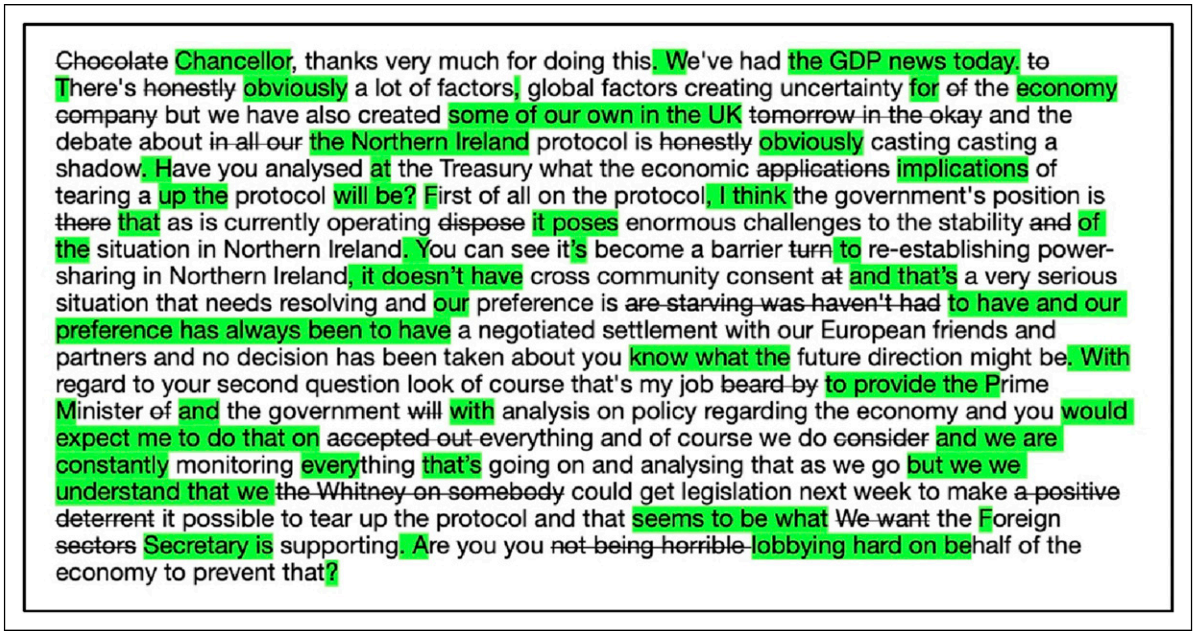
The following instructions apply to Windows 10 and may therefore not be applicable if using devices with newer or older operating systems. Please also note that, given that Word requires an Internet connection to transcribe, its use was not reviewed in this article. Hence, researchers transcribing whilst being connected to the Internet on a Windows device should check Microsoft's privacy policy prior to using it and check if there are any other steps that they should follow to ensure compliance with any data protection regulations in force in their country.

1. Configuring VB-CABLE. After downloading and installing VB-CABLE, you need to configure this as the system's audio output. To do so, click on the volume icon on the right-bottom corner of your screen > Open Sound Settings > Output and then select VB-CABLE. You will also need to configure it as the system's audio input. To do so, click on the volume icon on the right-bottom corner of your screen > Open Sound Settings > Input and then select VB-CABLE. Remember to restore the default output and input sources after you are done with transcription or else you will not be able to hear any sound from your device.

Appendix B

Using Pages to Transcribe Audio Files

Following recent scandals (Hern, 2021; see also Hern, 2019), Apple (2022b) allows users to process voice inputs directly on their device without therefore the need of having an Internet connection (see also Hern, 2021). In more detail, Apple (2022a, n. p.) explains that '[f]or general text Dictation performed on device [...], transcripts and audio are not shared with Apple by default, but are shared if you opt in to Improve Siri and Dictation'. Whether the Improve Siri and Dictation feature is turned



~~Chocolate~~ Chancellor, thanks very much for doing this. We've had the GDP news today. ~~to~~ There's ~~honestly~~ obviously a lot of factors, global factors creating uncertainty for ~~of~~ the economy ~~company~~ but we have also created some of our own in the UK ~~tomorrow in the okay~~ and the debate about ~~in all our~~ the Northern Ireland protocol is ~~honestly~~ obviously casting casting a shadow. Have you analysed at the Treasury what the economic ~~applications~~ implications of tearing a up the protocol will be? First of all on the protocol, I think the government's position is ~~there~~ that as is currently operating ~~dispose~~ it poses enormous challenges to the stability ~~and~~ of the situation in Northern Ireland. You can see it's become a barrier ~~turn~~ to re-establishing power-sharing in Northern Ireland, it doesn't have cross community consent ~~at~~ and that's a very serious situation that needs resolving and our preference is ~~are starving was haven't had~~ to have and our preference has always been to have a negotiated settlement with our European friends and partners and no decision has been taken about you know what the future direction might be. With regard to your second question look of course that's my job ~~beard by~~ to provide the Prime Minister ~~of~~ and the government ~~will~~ with analysis on policy regarding the economy and you would expect me to do that on ~~accepted out~~ everything and of course we do ~~consider~~ and we are constantly monitoring everything that's going on and analysing that as we go but we we understand that we ~~the Whitney on somebody~~ could get legislation next week to make ~~a positive deterrent~~ it possible to tear up the protocol and that seems to be what ~~We want~~ the Foreign ~~sectors~~ Secretary is supporting. Are you you ~~not being horrible~~ lobbying hard on behalf of the economy to prevent that?

Figure 4. The transcript generated offline by Pages on an Intel-based Apple laptop running Mojave of the first 90 seconds of Bloomberg (2022), after manual revision (author's own screenshot).

on can be found in the Keyboard Settings, which 'will indicate […] if your voice inputs and transcripts […] are processed on your device and not sent to or stored on Siri servers by default' (2022b, n. p.). However, it is worth noting that 'dictation requests […] are processed on the device' (Apple, n. d. a, n. p.) only if one is using more modern devices that come with an Apple silicon chip rather than Intel processors.

Therefore, researchers wishing to use Pages rather than MacWhisper to transcribe interviews fully offline as per the technique outlined in this article should use an Apple silicon device with at least Monterey installed[27] where Internet has been permanently disabled (purchasing a second-hand device might help minimise costs[28]). There is, however, an alternative. Until Catalina, Apple devices included indeed a feature called Enhanced Dictation (see Apple, n. d. d, n. d. e) which allowed users to dictate offline. Therefore, if in possession of, or considering buying, an older and less expensive Apple laptop with Intel processors rather than silicon chips,[29] one could still transcribe their interviews offline, provided that such a device has an operating system installed that comes with Enhanced Dictation, such as High Sierra or Mojave. It is worth noting that transcripts generated this way are less accurate (possibly because Enhanced Dictation relies on older language data) and punctuation needs to be added manually (see Figure 4 and further notes in Appendix A). Nonetheless, this could still be a helpful way to reduce transcription times, especially if one used a good quality recorder and the interviewee(s) spoke close to it. In either case, since the use of Pages was not reviewed in this article, researchers wishing to use it should consult a cybersecurity specialist before doing so.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fabio Battaglia https://orcid.org/0000-0002-3871-192X

Notes

1. https://www.nuance.com/dragon/business-solutions/dragon-professional.html [Accessed: 31/10/2023].
2. Da Silva (*ivi*, p. 9) used 'Arcam Solo Mini amplifier and Monitor Audio Silver RX2 speakers', both of which have been discontinued but which seem from an online search to have been sold at more than £600 each.
3. A male-to-male aux cable is a cable that has two jacks or pins at its ends, rather than a jack and a hole (male-to-female).
4. A three-ring male-to-male aux cable is an aux cable with three rings on both of its jacks that comes with a microphone channel.
5. https://vb-audio.com/Cable/.
6. https://www.audacityteam.org/.
7. https://www.speechpulse.com.
8. https://goodsnooze.gumroad.com/l/macwhisper.
9. Due to space constraints, this article focuses only on Windows and macOS, the two most used operating systems. The technique presented in this article was therefore not tested on other

operating systems such as Linux or ChromeOS, meaning that specific instructions or conclusions may not apply if using operating systems other than the ones reviewed.

10. Note that prices, as well as subscription plans, may change in the future. However, should these programmes' Premium plans become no longer affordable, or should their owners not offer lifetime subscriptions or free versions anymore, not only is it likely that other programmes will have been developed in the meantime to replace them (being Whisper open source), but as noted above there are ways to use Whisper models directly (see Stratvert, 2023).
11. Since they allow to upload and transcribe from audio files directly, both programmes can generate transcripts faster than if using live dictation whilst using aux cables or virtual audio devices, as to generate transcripts the latter way will take as long as the recordings themselves. However, in both cases this is time that the software and not the researcher requires to generate the transcript, hence since as I explain below transcription should be performed on an ad hoc device, this ultimately makes little difference as it is time that the researcher will dedicate to do other things anyway (as also noted by Da Silva, 2021).
12. Both programmes currently allow to translate interviews only into English. There are ways in the case of MacWhisper to expand the number of languages supported using DeepL, but '[b] e aware that when you translate your transcript, the content is sent to the DeepL servers' (MacWhisper, n. d. b, n. p.).
13. Some users have found ways to install it for offline use (see GitHub.com, 2023a) but doing so still requires some technical expertise.
14. Barnard et al. (2015), for instance, reported the case of someone who worked as a transcriber who claimed that it could take them up to 14 hours.
15. This figure was estimated considering a total number of 120 hours of conversation and a rough average cost for transcribing interviews manually in the United Kingdom of £1 per minute – this being lower than the starting price per minute charged by the company recommended by my university (£1.35).
16. Recordings can be played directly from the recorder (see Picture 1) or from another device where they have been copied onto (and encrypted), using a three-ring aux cable as explained above. Alternatively, after having been encrypted they could be transferred via a Universal Storage Bus (USB) drive or a Secure Digital (SD) card and, once ready, (encrypted) transcripts could be transferred to one's main device with the same or another USB drive or SD card.
17. When turning the Internet off, a good idea is to delete all network profiles so that the device does not automatically connect to them should the Internet be switched on by mistake. For further protection, in addition to disabling the Wi-Fi, any Ethernet cables should be unplugged, too, and Bluetooth turned off.
18. In theory, this is not needed if transcripts are generated offline on a device that is kept offline. However, it is always a good idea to take extra security measures which could prevent e.g. data from being sent should the Internet be switched on by mistake.
19. See SpeechPulse (2023a) and MacWhisper (2023).
20. At the time of writing (October 2023), second-hand Apple devices that support Ventura (which is required to install MacWhisper) can be purchased on eBay.co.uk from ca. £400. If transcribing more than ca. ten hours of interviews in an entire lifetime, this is going to be cheaper than purchasing other voice recognition programmes, subscribing to transcription platforms or hiring professional transcribers (see prices mentioned throughout the article). Second-hand Windows devices that meet the minimum requirements can be purchased instead for even less, with prices starting from ca. £200/300. If purchasing a second-hand device, make sure that this is reset to factory settings (see footnote no. 22) and that it is safe to use (see Da Silva, 2021), consulting a professional if necessary.
21. Please note, due to lack of funding it was not possible for me to test either SpeechPulse or MacWhisper on a tablet.
22. Since data can be recovered even after a factory reset, it is important to perform this with the advice of a professional in order to ensure that data have been encrypted and deleted permanently.
23. https://www.microsoft.com/en-gb/microsoft-365/buy/microsoft-365 [Accessed: 07/09/2023].
24. While some of these platforms offer the possibility of setting up an account with a limited number of free transcription hours (e.g. Descript), it shall not be forgotten that these are businesses like any other and therefore require the purchase of a product, often in the form of a subscription plan. In some cases, such plans can be relatively economical: Descript's Creator plan, for instance, allows to transcribe 100 hours of audio in a year for $144 (https://www.descript.com/pricing; accessed: 31/10/2023). To transcribe the same number of hours with Amberscript's pre-paid plan, instead, one would have to pay €1,000 – Amberscript also offers cheaper annual plans, but these only allow to transcribe 5 hours of audio or video uploaded per month (https://www.amberscript.com/en/pricing/; accessed: 31/10/2023). To transcribe 1,000 hours of audio offline with Simon Says's one would need instead to pay $10,000 (https://www.simonsaysai.com/buy-on-premises-ai-transcription-solution; accessed: 31/10/2023).
25. I contacted Nuance, the company owning Dragon, five times for some clarifications about their privacy policy but they never responded. Otter.ai, too, was approached for some clarifications but never responded.
26. From SpeechTexter's (2021, n. p.) Privacy Policy: 'SpeechTexter does not store any text you dictate. All speech is processed at Google servers, they carry their own privacy policy'. From LilySpeech's (2020, n. p.) Privacy Policy: 'None of your dictation passes through our servers. The actual speech-to-text conversion is handled by Google'.
27. This is because in Big Sur 'your dictated utterances are sent to Apple' (Apple, n. d. b, n. p.) and in Catalina, too, 'your words are sent to Apple servers' (Apple, n. d. c, n. p.).
28. See, however, Da Silva (2021).
29. At the time of writing (October 2023), older Intel-based Apple laptops can be found on eBay.co.uk for ca. 100£.

References

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., Bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al

Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy: RSAP*, *19*(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Apple. (2022a). *Improve Siri and dictation & privacy*. Apple. https://www.apple.com/legal/privacy/data/en/improve-siri-dictation/

Apple. (2022b). *Ask Siri, dictation & privacy*. Apple. https://www.apple.com/legal/privacy/data/en/ask-siri-dictation/

Apple-a. n. d. a. *Dictate messages and documents on Mac*. Apple. https://support.apple.com/en-gb/guide/mac-help/mh40584/13.0/mac/13.0

Apple-b. n. d. b. *Dictate messages and documents on Mac*. Apple. https://support.apple.com/en-gb/guide/mac-help/mh40584/11.0/mac/11.0

Apple-c. n. d. c. *Dictate your messages and documents on Mac*. Apple. https://support.apple.com/en-gb/guide/mac-help/mh40584/10.15/mac/10.15

Apple-d. n. d. d. *Dictate your messages and documents on Mac*. Apple. https://support.apple.com/en-gb/guide/mac-help/mh40584/10.14/mac/10.14

Apple-e. n. d. e. *Dictate your messages and documents*. Apple. https://support.apple.com/en-gb/guide/mac-help/mh40584/10.13/mac/10.13

Bailey, C. A. (2023). *A guide to qualitative field research*. Sage Publications Ltd.

Barnard, H., Hjorth, I., Graham, M., & Lehdonvirta, V. (2015). Online labour markets and the persistence of personal networks: Evidence from workers in Southeast Asia. In Paper Presented at the American Sociological Association Annual Meeting 2015, Chicago, USA, 2015. https://vili.lehdonvirta.com/files/OnlinelabourmarketsandpersonalnetworksASA2015.pdf

Bloomberg. (2022). *UK chancellor of the exchequer Rishi Sunak full interview*. Bloomberg. https://www.youtube.com/watch?v=O2M7-PTdt4Q

Bogdan, R., DeVault, M. L., & Taylor, S. J. (2016). *Introduction to qualitative research methods. A guidebook and resource* (4th ed.). John Wiley & Sons, Inc.

Bokhove, C., & Downey, C. (2018). Automated generation of 'good enough' transcripts as a first step to transcription of audio-recorded data. *Methodological Innovations*, *11*(2), 1–14. https://doi.org/10.1177/2059799118790743

Bolden, G. B. (2015). Transcribing *as* Research: "Manual" Transcription and Conversation Analysis. *Research on Language and Social Interaction*, *48*(3), 276–280.

Bourgeault, I., & Corrente, M. (2022). *Innovation in transcribing data: Meet Otter.ai* (pp. 1–27). Sage Research Methods: Doing Research Online.

Brinkmann, S. (2022). *Qualitative interviewing: Conversational knowledge through research interviews* (2nd ed.). Oxford University Press.

Brinkmann, S., & Kvale, S. (2018). *Doing Interviews* (2nd ed.). Sage Publications Ltd.

Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, *8*(25), 1–11. https://doi.org/10.1057/s41599-020-00703-8

Chubb, J., Cowling, P., & Reed, D. (2022). Speeding up to keep up: Exploring the use of AI in the research process. *AI & Society*, *37*(4), 1439–1457. https://doi.org/10.1007/s00146-021-01259-0

Da Silva, J. (2021). Producing 'good enough' automated transcripts securely: Extending Bokhove and Downey (2018) to address security concerns. *Methodological Innovations*, *14*(1), 1–11. https://doi.org/10.1177/2059799120987766

De Felice, D., & Janesick, V. J. (2015). Understanding the marriage of technology and phenomenological research: From design to analysis. *Qualitative Report*, *20*(10), 1576–1593. https://doi.org/10.46743/2160-3715/2015.2326

Dempster, P. G., Lester, J. N., & Paulus, T. M. (2015). *Digital tools for qualitative research*. Sage Publications Ltd.

di Gregorio, S. (2022). Voice to text: Automating transcription. In P. Mihas, J. Saldaña, & C. Vanover (Eds.), *Analyzing and interpreting qualitative research: After the interview* (pp. 97–112). Sage Publications, Inc.

Erickson, T., Voida, A., Mynatt, E. D., & Kellogg, W. A. (2004). Interviewing over instant messaging. In *CHI EA '04: CHI '04 extended abstracts on human factors in computing systems* (pp. 1344–1347). Association for Computing Machinery.

Fletcher, A. K., & Shaw, G. (2011). How voice-recognition software presents a useful transcription tool for qualitative and mixed methods researchers. *International Journal of Multiple Research Approaches*, *5*(2), 200–206. https://doi.org/10.5172/mra.2011.5.2.200

Flick, U. (2018). *An introduction to qualitative research* (6th ed.). Sage Publications Ltd.

GitHub.com. (2023a). *How to install and use Whisper offline (no internet required) #1463*. GitHub.com. https://github.com/openai/whisper/discussions/1463

GitHub.com. (2023b). *Openai/whisper*. GitHub.com. https://github.com/openai/whisper/blob/main/README.md

Hern, A. (2019). *Apple contractors 'regularly hear confidential details' on Siri recordings*. The Guardian. https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

Hern, A. (2021). *Apple overhauls Siri to address privacy concerns and improve performance*. The Guardian. https://www.theguardian.com/technology/2021/jun/07/apple-overhauls-siri-to-address-privacy-concerns-and-improve-performance

Hest, J. (2022). An unexpected journey: From typing to dictating a thesis. *Waikato Journal of Education*, *27*(2), 91–95. https://doi.org/10.15663/wje.v27i2.929

Horrocks, C., & King, N. (2010). *Interviews in Qualitative Research*. SAGE Publications Inc.

Jarnow, J. (2016). *Why our crazy-smart AI still sucks at transcribing speech*. Wired. https://www.wired.com/2016/04/long-form-voice-transcription/

Johnson, B. E. (2011). The speed and accuracy of voice recognition software-assisted transcription versus the listen-and-type method: A research note. *Qualitative Research*, *11*(1), 91–97. https://doi.org/10.1177/1468794110385966

Kawahara, T. (2012). Transcription system using automatic speech recognition for the Japanese parliament (Diet). In Proceedings of the Twenty-Fourth Innovative Applications of Artificial

Intelligence Conference, Toronto, Canada, July 22–26, 2012. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ca66f0725ee52026ca838c94bc5771818e801086

Kawahara, T., Mimura, M., & Sakai, S. (2021). An end-to-end model from speech to clean transcript for parliamentary meetings. In Proceedings of the APSIPA Annual Summit and Conference, Tokyo, Japan, December 14–17, 2021. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9689457

Kine, P. (2022). *My journey down the rabbit hole of every journalist's favorite app*. Politico. https://www.politico.com/news/2022/02/16/my-journey-down-the-rabbit-hole-of-every-journalists-favorite-app-00009216

LilySpeech. (2020). *Terms of service*. LilySpeech. https://lilyspeech.com/terms-of-service/

MacLean, L. M., Meyer, M., & Estable, A. (2004). Improving accuracy of transcripts in qualitative research. *Qualitative Health Research*, *14*(1), 113–123. https://doi.org/10.1177/1049732303259804

MacWhisper. (2023). *Whisper transcription terms & privacy policy*. MacWhisper. https://impresskit.net/press-release/f5537c9f-c2c1-42d9-a870-6cecd28a8c31

MacWhisper-a. nd a. *MacWhisper*. MacWhisper. https://goodsnooze.gumroad.com/l/macwhisper

MacWhisper-b. nd b. *MacWhisper release note*. MacWhisper. https://macwhisper-site.vercel.app/releases/release_notes.html

Matheson, J. L. (2007). The voice transcription technique: Use of voice recognition software to transcribe digital interview data in qualitative research. *Qualitative Report*, *12*(4), 547–560.

McCartan, K., & Robson, C. (2016). *Real world research* (4th ed.). Wiley-Blackwell.

McMullin, C. (2021). Transcription and qualitative methods: Implications for third sector research. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *34*(1), 140–153. https://doi.org/10.1007/s11266-021-00400-3

Morris, A. (2015). *A practical introduction to in-depth interviewing*. Sage Publications Ltd.

Park, J., & Zeanah, A. E. (2005). An evaluation of voice recognition software for use in interview-based research: A research note. *Qualitative Research*, *5*(2), 245–251. https://doi.org/10.1177/1468794105050837

Point, S., & Baruch, Y. (2023). (Re)thinking transcription strategies: Current challenges and future research directions. *Scandinavian Journal of Management*, *39*(2), 1–10. https://doi.org/10.1016/j.scaman.2023.101272

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). *Robust speech recognition via large-scale weak supervision*. In A. Anders, A. Alexandr, Y. C. Justin, I. Piotr, N. Shyam, & S. Sandeep (Eds.), Proceedings of the 40th International Conference on Machine Learning. PMLR

Richardson, J., Godfrey, B., & Walklate, S. (2021). Rapid, remote and responsive research during COVID-19. *Methodological Innovations*, *14*(1), 1–9. https://doi.org/10.1177/20597991211008581

Sarkar, S. (2021). Using qualitative approaches in the era of big data: A confessional tale of a behavioral researcher. *Journal of Information Technology Case and Application Research*, *23*(2), 139–144. https://doi.org/10.1080/15228053.2021.1916229

Shelton, M., & Grauer, Y. (2022). *How secure are journalists' favorite transcription tools?* Freedom of the Press Foundation. https://freedom.press/training/blog/how-secure-are-journalists-favorite-transcription-tools/.

Silver, C., & Lewins, A. F. (2020). Computer-Assisted analysis of qualitative research. In P. Leavy (Ed.), *The oxford handbook of qualitative research* (2nd ed., pp. 912–955). Oxford University.

SpeechPulse. (2023a). *Download SpeechPulse*. SpeechPulse. https://speechpulse.com/download/

SpeechPulse. (2023b). *Voice typing everywhere*. SpeechPulse. https://speechpulse.com/

SpeechTexter. (2021). *Privacy policy*. SpeechTexter. https://www.speechtexter.com/privacy

Stratvert, K. (2023). *How to install & use whisper AI voice to text*. Kevin Stratvert. https://www.youtube.com/watch?v=ABFqbY_rmEk.

Tang, R. (2023). Manual transcription. In J. M. Okoko, S. Tunison, & K. D. Walker (Eds.), *Varieties of qualitative research methods* (pp. 295–302). Springer.

Walker, M. (2006). *Southern farmers and their stories: Memory and meaning in oral history*. The University Press of Kentucky.