Covid-19 and Interview Mode Debates: Reflections on Using WhatsApp for Voice-only Interviewing

MUHAMMAD SALMAN KHAN
M.SALMAN.K85@GMAIL.COM

KEYWORDS: QUALITATIVE RESEARCH, VOICE-ONLY INTERVIEWS, WHATSAPP, ONLINE INTERVIEW MODES, GENDER, SWAT VALLEY

Introduction

Qualitative researchers in social science are debating the use of face-to-face and online interviewing for over three decades (Obdenakker 2006; Irvine et al. 2011; Johnson et al. 2019). This debate spans across disciplines including health sciences (Pell et al. 2019), management studies (Farooq et al. 2017), social work (Holt 2010), and social science more broadly (Johnson et al. 2019). In the pre-pandemic world, some qualitative research textbooks did not pay attention to online interviews at all, while others regarded them as 'a last resort' (Hermanowicz 2002: 497), 'not a preferred way' (Rubin & Rubin 2011: 125), or had generally second-rated them in qualitative research (Holt 2010). While proponents consider them as acceptable and necessary (Cachia & Millward 2011), 'equally good for some research contexts' (Farooq et al. 2017), or even better for some sensitive topics (Sturges & Hanrahan 2004), criticism of online modes of interviewing (Johnson et al. 2019) is centred around their disembodied character1 lack of visual context of interview setting, absence of visual cues to guide the conversation, and accessibility of participants to technology such as telephones, smartphones, and computers. This entire range of debates around the modes of interviewing consists both of reflective accounts and studies specifically designed to compare face-to-face interviews with different modes of
online interviewing such as email (James & Busher 2007), telephone (Holt 2010), videoconferencing or skype (Janghorban et al. 2014; Obdenakker 2006).

Contributing to this rapidly growing literature, the present paper is a reflective account of using online modes of remote interviewing in the post-Covid research landscape where online interviewing was not a choice but a necessity imposed by Covid-19 (Lawrence 2020). Using technology for online interviewing is geographic in nature that requires ethical and practical considerations. Important among these considerations include country-specific institutional regulations related to online platforms, lack of safety of certain online platforms such as Zoom, interviewee’s preference for certain modes (such as voice-over video), and researcher’s ability to interview some participants via certain modes (Lawrence 2020; Mwambari et al. 2021: 4; O’Connor et al. 2008). After considering these factors in the context of Pakistan, this study relied on WhatsApp to conduct voice-only interviews (voice-only interviews hereafter).

This paper will not repeat the well-rehearsed arguments on both the sides of the mode-effect (video or voice-only) debates on the quality of interviews in terms of their word-count, duration and capacity to produce topic-related data (Vogl 2013; Sterges & Hanrahan 2004; Johnson et al. 2019). Instead, four dimensions of voice-only interviewing are analysed: (1) Providing access to harder-to-reach participants, (2) The significance of culturally sensitive listening in the absence of visual context of the interview setting, (3) Increased power to the interviewees within the power dynamics of the researcher-participant relation, and (4) The advantages of using WhatsApp over traditional telephone connections for qualitative interviews. Reflections offered in this paper contribute not only to the research methods debates over the modes of interviews, but also to social science research in South Asia and particularly to the scholarship on Pakistan-Afghanistan borderlands.

With a few exceptions, most of the literature in the research methods debates over the modes of interviews is situated in the Global North contexts (Johnson et al. 2019). Covid-19 forced researchers to adopt online methods for their researches around the world, but in the post-Covid methods literature on Global South, researcher’s positionality in terms of gender and its implications for voice-only (and online more broadly) interviewing is overlooked (Hall et al. 2021). Reflections on positionality do exist, but mostly in researches by women about women, and by men about men. Khyber Pakhtunkhwa province of Pakistan is an example of this methodological trend in the pre-Covid offline, and post-Covid online social science research. For instance, pre-Covid qualitative
studies in anthropology, sociology, political science, and business and economics in the context of the Khyber Pakhtunkhwa province of Pakistan (where our fieldwork was conducted) justify the gendered composition of their sample on the grounds of local gendered norms as an impediment to recruiting participants from the opposite sex (Ahmed 1980; Lindholm 1982; Saeed 2012 among others). For the purposes of this paper, as a male researcher I classify Swati women as hard-to-reach participants because they are difficult-to-access due to cultural norms of gender segregation, and they are hidden because no reliable records of their online presence exist (Ellard-Gray et al. 2015: 1f.). Local cultural norms not only impact researcher’s ability to recruit participants from opposite gender, they also govern the use of technology by men and women (Ibtasam et al. 2019).

Our research field is the Swat valley (Swat hereafter) in the Khyber Pakhtunkhwa province of Pakistan. Swat has a rich cultural and political history and it is a historic centre of Buddhist, Hindu, and Muslim civilisations. Until 1969, when Swat merged with the Pakistani state, it remained a princely state with its distinct governance and development apparatus (Sultan-e-Room 2008). In the twenty first century, specifically between 2007 and 2009, Swat became the centre of the Tehreek-i Taliban Pakistan’s violence that led to a military operation against the Taliban. This military operation resulted in the largest crises of internally displaced population. More than 2 million people fled the region, large parts of government infrastructure (especially girls' schools) were destroyed (ICG 2013), and historical monuments (such as the famous Jihanabad Buddha statue) were defaced (De Nardi 2018). The spaces associated with women’s socialisation and professional engagement such as women’s markets, springs and Shrines (traditional sites of socialisation), girls' schools, NGOs, and dancing spaces came under the Taliban’s attack (Khan 2021). The Taliban’s violence, and the military offences to expel them 'were fear-inspiring' for many women inhabitants of Swat (Marsden & Hopkins 2013: 8). In the post-Taliban Swat, our emerging findings suggest, women’s mobility has increased however, interactions between men and women are still restricted by the Pakhtun cultural norms of women’s seclusion and family’s honour. These restrictions have historically limited access of men researchers to women participants in this region.

Within this context, the reflections offered in this paper on voice-only online interviewing contribute to methodological advancement both in terms of the gendered research geography of the Pakistan-Afghanistan borderland, and in qualitative methods debates over the means of remote interviewing. Telephone is not the only remote interviewing mode, other remote interviewing formats have also attracted ample
attention in the research methods literature. For instance, web-based video conferencing tools such as Skype are also highlighted by scholars for their potential benefit of making the interview interaction 'as comparable to the onsite equivalent for the presence of nonverbal and social cues' (Janghorban et al. 2014: 1). Others point out the technical problems such as dropped calls, poor audio quality or frozen videos due to poor internet connections, and inability of the researchers to view the full body-language of the interviewee depending on the position of the camera (Seitz 2016). The significance of visual cues for an interview depends upon the interview sample, and the research questions (Farooq et al. 2017). I was aware of these considerations; however, the choice of mode between voice-only and video interviewing was not guided by these technical considerations alone. Instead, cultural norms of women’s seclusion, interviewee’s preference for voice-only over video interview, digital competency (and comfort with technology) of the women users in Pakistan, and other practicalities associated with available technologies made WhatsApp as the preferred medium of online interviewing (for details see methods section).

Against this backdrop, this paper asks: How do synchronous (in time but asynchronous in place) voice-only interviews facilitate access to hard-to-reach research participants? How can they benefit from culturally sensitive listening in the absence of visual access to the interview settings of the interviewees? In what ways do voice-only interview empower interviewees in the interview process and how that empowerment is beneficial for the interviews? How WhatsApp, as a technology for online interviewing, fits better in the lives of people of Swat Valley of Pakistan. It must be noted that methodologically, WhatsApp interviews have certain advantages over telephone interviews which are pointed out in the methods and penultimate sections (section 2 and 6) of this paper.

In what follows, section two outlines the methodology including methodological reflections, sample size and characteristics, and method of conducting interviews including ethical considerations. Section three demonstrates how voice-only interviews (whether telephonic or WhatsApp) interviews facilitate access to hard-to-reach population in the culturally specific, gendered-segregated context of Swat Valley. Section four reports findings on how culturally sensitive listening can improve the quality of voice-only interviews in the absence of visual cues. Section five explores the ways in which I found online interviews as empowering for research participants. Section six reflects upon the advantages and caveats of conducting voice-only interviews via WhatsApp compared to telephone. Section seven concludes.
Methodology

This paper is a reflective account of my involvement in an ongoing project, which was designed and started in 2019 within the pre-pandemic world. This project is interested in learning how the communities and landscapes in Swat are healing from, and reconciling with the wounds inflicted by the Taliban violence, more than a decade ago (Marsden & Hopkins 2013). The project, titled "Mapping the wounded landscape of Swat Valley: From Taliban violence and iconoclasm to social growth and reconciliation" has four thematic dimensions including poetry, historical heritage (archaeological) sites, and lived heritage including women's markets, and natural resources such as springs, forests and orchards. The project is a collaborative endeavour involving qualitative researchers from King’s College London and the School of Oriental and African Studies, based in London/United Kingdom, as well as from Western Sydney University in Australia along with the University of Malakand, based in Chakdara/Pakistan as local partner.

As the only Postdoctoral Research Associate on the project, I am responsible for liaising with all the partners, and mentoring and working closely with two research assistants (RAs henceforth). The two RAs are Mohsin Ullah (male RA) and Aneela Gul (female RA) who are employed for one year to assist in conducting field research. Mohsin Ullah, who is a doctoral student at Quaid-i-Azam University (Pakistan) and Aneela Gul who is a development practitioner are not just research assistants, but junior colleagues who enjoy hands-on training and full mentorship support. Key elements of their training include being reflexive, ability to write and publish under their own names, and conduct in-depth interviews. This entails regular fortnightly meetings, debriefing after every interview, and no harm to participants and themselves (RAs). Reflexivity and continuous debriefing allowed me to compare Aneela’s face-to-face interviews with women in Swat and my online, voice-only interviews with women (which is the central concern of this paper).

My complex positionality as a culturally embedded, local male researcher working in Western academia afforded me an opportunity to develop a reflexive account of interviewing women through voice-only interviews. For me, justifying absence of women from qualitative sampling frames on the grounds of gender by native male researchers as a methodological limitation has always been a sign of epistemic lethargy (Khan 2019: Ch4). It is not to argue that my positionality as a male researcher as a hindrance to access women participants was mitigated by my status as member of a local family working in Western academia. Instead, awareness of, sensitivity to and respect for local culture, non-kin local social networks, and openness to serendipitously
arising opportunities as opposed to ex-post closure of sampling frame can help mitigate (to a considerable extent) gender norms as a barrier to recruiting women participants. Our (my team’s) project design initially aimed to recruit a local female RA to conduct interviews with women in the region. However, the possibility of interviewing women through voice-only interviews emerged serendipitously before the recruitment of research assistants. After analysing textual and audio-visual material on conflict-related poetry, my colleague Dr. James Caron and I decided to conduct online interviews with poets whose books were analysed in greater depth (including two women authors). From a basement flat in Stockwell, South London, my personal networks and WhatsApp were the only means to reach out to these potential interviewees. My personal acquaintances in the region were helpful in providing contact numbers of male poets who were directly called for requesting an interview at a future date. Whereas, for women poets, their colleagues who were my acquaintances had negotiated access.

Like most qualitative researches, my project team began with the "gold-standard" assumption that interviews will be conducted in-person (Johnson et al. 2019). Online modes of interviewing were not a consideration at all in the initial design of our project. All four team members based in western academic institutions (including myself) have extensive fieldwork experience in the Northwest of Pakistan. For their respective thematic focus, each was bound to visit Pakistan for conducting fieldwork for this project. Travel limitations and health risks posed by Covid-19 to both the researchers and research participants forced us to rethink our research design and incorporate remote interviewing into it (Sy et al. 2020: 602f.). To keep the project up and running amid the uncertainty posed by Covid-19 on the possibility of in-person interviewing, my project colleagues turned to online methods. Email interviews were initially tried with the possibility of complimenting them with face-to-face interviews at later stages (James & Busher 2006).

Dr. Sarah De Nardi, an international expert on community-led heritage practices with extensive networks of collaborators in Islamabad and Swat was the first in our team to attempt an email interview with a journalist and a local heritage activist in Swat, which was a success. The interview was in English because Dr Sarah does not speak Pashto and the interviewee is fluent in English. However, when I conducted a follow-up interview with the same interviewee in Pashto over WhatsApp, he referred to the email interview and stated, 'I [interviewee] think if the same [email] interview was conducted in Pashto [first language] or in Urdu [second language] I could have responded better and of course in more detail' (4 June 2020). This comment of our interviewee, who has published extensively in English, was enough to rule out email as
asynchronous interviews and consider voice-only synchronous interviews in Pashto as the only choice.

Email interviews could also be conducted in Pashto, but in Swat, neither email is the preferred means of communication nor did most of our interviewees have the technical proficiency to use Pashto alphabets on an English keyboard. In the first thematic focus on interviews with poets, I co-conducted (along with my colleague Dr James Caron) 13 remote interviews with 11 male and two female poets in July and August 2020. In four of the interviews with our 11 male participants, video interviews had to be turned into voice interviews to improve the flow of interview that was affected by unstable internet connections. Female interviewees accepted our interview requests only if the interview was conducted without video appearance. Telephone interviews were not employed for conducting voice calls due to financial costs involved for telephone calls between London/UK and Swat/Pakistan. Zoom was deliberately avoided because of concerns over its security and potential consequences for our participants in a conflict-affected region (Mwambari et al. 2019). Although Signal has enhanced security than WhatsApp, we did not ask participants to install it because asking participant to install an app with additional security features might have generated distrust regarding our project aims. Skype as an option was discarded because the interviewees already had WhatsApp. Moreover, unlike Skype, WhatsApp is available on feature phones, which makes WhatsApp more widely accessible in Pakistan. Therefore, for the thematic focus on women’s markets and their affective dimension, voice-only interviews were conducted over WhatsApp because it involved women interviewees only.

I have conducted 10 remote interviews at this stage between November 2020 and March 2021, along with six face-to-face interviews conducted by Aneela who is located in the study area (see Appendix 1).¹⁵ None of the interviews were audio-recorded, due to the concern that a recording might affect the quality of the data. Interviewees might be less comfortable and more formal when recorded, or even reluctant to be interviewed (Al-Yateem 2012). Audio/electronic recording may also create distrust in a context where people have concerns about the cause of data generation, why the interviewee is being recorded, and where the data will be published (Hays & Mattimoe 2004). In the conflict-affected Swat Valley of Pakistan, this may especially hinder recruitment of interviewees and their openness to share information. We (project team members) decided to stick to this commitment originally proposed in our research ethics application for face-to-face interviews. Therefore, the interviewers relied on context notes to describe occasional nonvisual characteristics of the interview, such as nonverbal characteristics of the
respondent or the setting in which the interview was situated (when relevant) (Johnson et al. 2019: 5f.).

The voice-only interviews with women lasted between 11 minutes to one hour. The reason for the shortest interview was not the mode but the interview setting which will be explained in the next section. All the interviewees were recruited through snowball sampling through their acquaintances, and once they agreed to participate, their verbal consent was obtained. The information sheet and consent form were translated into both Urdu and Pashto, recorded using the voice-note function in WhatsApp to be sent to each interviewee individually. After obtaining the verbal consent, the researcher signed the form as an affirmation that the consent was indeed obtained. These consent forms along with the business cards of the researcher, and contact number of the University of Malakand (our local partner) was provided so that the participants can ask for withdrawal from the research at any time before 31 August 2021. However, our first interviewee reported non-availability of the PDF or word readers in their phones. In such instances, physical copies of the form were sent to the male participants using my local contacts, whereas none of the women interviewees reported difficulties with receiving electronic copies.

All the interviews were conducted in Pashto, which is my native language as well as of all the interviewees. The first seven interviewees were not asked questions about their interviewing experience, reflective notes on contact negotiation and context notes on the interview dynamics are analysed in this paper. From the eighth interview onwards, the closing question was regarding the interviewee’s experience of voice-only interviewing. Invariably all the interviews offered interesting non-visual, specific to each interview setting, background cues. In some instances, notes on these background non-visual cues are also compared with reflective notes of Aneela who conducted six face-to-face interviews in Swat in November-December 2020. In the following, I analyse my reflective notes on contact negotiation and context notes on the nonvisual dynamics of interview settings to produce the next three sections. These sections contextualise my findings in the broader methodological debates over voice-only interviews.

**Voice-only interviews facilitate accessing hard-to-reach population**

Telephone interviews are said to reduce costs associated with face-to-face interviews especially when interviewees are geographically dispersed (Irvine et al. 2011). In addition, due to synchronous communication in time but asynchronous communication in place, telephone interviews
have the advantage of providing extended access to participants, compared to face-to-face interviewing (Obdenacker 2006). Mann and Stewart (2000) note, among other advantages, that telephone interviews enable researchers to contact hard-to-access populations, particularly for face-to-face interviewing, for instance, mothers at home with small children and people with disabilities. These asynchronous place interviews can facilitate access to sites that have closed or limited access, for instance, hospitals, religious communities, the military, cults and prisons. Telephone interviews are also helpful in accessing dangerous and politically sensitive sites such as war zones or sites where viruses are rife, without needing to grapple with the danger of virus and the bureaucratic challenges for visiting the sites (Obdenakker 2006). In our case, access to avoid the danger of virus (Covid-19) and related bureaucratic hurdles (such as risk assessments and international travel bans) was the starting point for conducting interviews over telephone.

Research consistently shows that in Pakistan, mobile phones are the dominant medium of accessing social media and online communication platforms for a great majority of women internet users. However, access to technology is socially mediated (Ibtasam et al. 2019). Families consider women’s access to mobile phones bad, which limit their possibilities to have access to online communication technologies. Family’s surveillance, hate speech, harassment and privacy breach in the online sphere are the factors limiting women’s access to communication technology and using it to converse with strangers (Aksar et al. 2020: 993f.). Although the need to use internet during Covid-19 has increased, women are facing increased difficulties in accessing and using the internet. The ability to use internet outside their homes where family surveillance can be avoided, such as educational institutions, workplaces and libraries was reduced due to lockdown and mobility restriction that had been put in place due to the coronavirus (Rehman et al. 2021: 6). Thus, for engaging women in in-depth voice-only (online) interviews, family’s surveillance, interview setting (home or workplace), phone ownership (own or shared), and negotiating access (by the researcher as a stranger or through the interviewee’s trusted contact) should be key considerations in the context of Pakistan. Online researchers need to reject viewing Pakhtun (or Pakistani) women as a monolithic category, either inaccessible to male researchers (Jan 2010; Saeed 2012), or representing uniformed gendered voice without aired by class, education, profession, age, and individual idiosyncrasies (cf. Weiss 2013: 180-90).

I was quick to experience that voice-only interviewing holds the promise of accessing women interviewees who were previously inaccessible to male researchers for face-to-face interviews (Ahmed 1980; Jan
2010; Saeed 2012). However, the considerations related to culturally mediated presence of women in the online world noted in the previous paragraph, guided the iteration of our research design in response to Covid-19.

Not only that, the first two interviews with women poets revealed that women interviewees wanted their perspectives to be heard and were more forthcoming in providing information on their experiences related to the regional conflict (our research topic). This finding was cautiously received to avoid any generalisations. Initially, it was assumed that these two interviewees are poets and educated professional women which might be the reason for them wanting to be heard. Additional interviews with women of different ages and professions confirmed Weiss (2012: 180) experience that Swati women want their perspective to be heard.

Researchers must pay special attention to the social norms of women’s seclusion, women’s limited ability to talk to male strangers over the phone, and strict family surveillance over their phone calls. Within this context, random recruitment does not work and may be a significant hurdle to recruit participants in the cultural geographic context where this research was conducted. In the first place, records of women as telephone owners do not exist. Secondly, women either do not respond at all, or do not respond positively to calls from strangers. Third, as also reported in the male researcher’s vignette (Rahat Shah in this issue), interviewers may have to negotiate access with a male family member first who is often unwilling for his female family member to be interviewed by a strange man. Even if in a very exceptional case, the male family member of a potential research participant allows her to be interviewed by a male researcher, the interviewee will not express her thoughts independently because of the presence of that male family member in the interview setting.

To avoid these potential negative influences on the quality of telephone interviews, randomly accessing women for interviews should be avoided. Access in our case was negotiated by local men and women who were acquainted with our potential interviewees. Men were helpful in negotiating access in the professional settings with professional women only, whereas only women were helpful in negotiating access to non-working women. None of our interviewees for voice-only interviews was illiterate (see Appendix 1). Therefore, our claim that telephone interview facilitate access to women as hard-to-reach participants for male researchers where face-to-face interviews are difficult do not extend to uneducated women for whom we have no evidence.
Is culturally sensitive listening a pre-requisite for distant, voice-only interviewing?

Relying on voice-only interview generated a greater sense of powerlessness for me, as the interviewer in the interview setting. As the interviewer I was deprived of any embodied experience whether the interviewee was surrounded by people or was alone, surrounded by male or female family members etc. In a face-to-face interview, presence of other people in the interview setting and interviewee’s nonverbal communication with them while responding, space available to the interviewee (private room, living room, courtyard at home, or shared/personal workplace), interviewee’s body language exhibiting concern of being overheard, all are important background visual cues for contextualizing interviewee’s responses.

Regardless of the mode (either face-to-face or voice-only online), even if a female researcher is interviewing women, the likely presence of a male family member often also makes the interview setting uncomfortable because the male family member might respond on interviewee’s behalf, dismiss interviewer’s questions, or constantly stare at the interviewer and interviewee (Ibtasam et al. 2020: 11). If an interviewee feels that a male family member might overhear her, the interview may temporarily or permanently be disrupted even if that member is not interrupting the interview conversation. Reflecting upon a face-to-face interview with a 25-year-old woman about her experiences in the women’s markets, our research assistant observes:

When I asked her [interviewee] the question, instead of answering, she looked towards the main door [of the house], because her brother came inside; he went to the kitchen which was in the lawn [where we were sitting]. 'I understood that she feels shy talking in front of him'. (Aneela’s interview notes, 14 Dec. 2020)

To avoid such a situation in a voice-only interview with me (as a male researcher), one of my interviewees preferred her office hours stating: "I can talk in a more relaxed environment there [in the office] because no one from my family will be listening to what I am saying about my experiences in the market" (field note, 3 Dec. 2020). In those instances, where home was the only option for the interviewee in terms of settings/timings, I relied on auditory cues in the background to comprehend the specific setting and contextualised her responses instead of blaming the mode for short responses or non-richness of the interview (Johnson et al. 2019).

Age or profession of an interviewee are not the bases of this finding. One may also be tempted to think that the degree of acquaintance of the lead (male) or associate (female) researcher with the participant
might have affected the quality and length of responses in an online interview. Undoubtedly, these factors are important, for instance, the interviews arranged for me by Aneela with some of her ex-colleagues working in various NGOs were long (in duration) and rich (in details) than the interviews negotiated by my personal contacts. However, two interviews in particular, with a 50-year-old female poet and a 23-year-old Master’s student lasted even longer than those negotiated by Aneela. It was my first-ever conversation with both the poet (age 50) and the student (age 23). I asked a male friend (also a poet) to contact the poet after reading her book, while another female respondent who was the interviewee’s classmate contacted the second. Both of the interviewees (the poet and the student) were at home at the time of their interviews.

The absence of a visual background for the interview setting re-enforced greater reliance on listening as opposed to visual sensory experiences. At the beginning of my interview with the 23-year-old Master’s student mentioned above, the interviewee sounded cheerful and confident until a sudden knock on the door after a few introductory sentences re-oriented this atmosphere. As soon as the interviewee heard the knocking, without taking a microsecond, she asked me to excuse her so that she can attend to the door. The sudden change in the interviewee’s tone in response to the knocking at her door deprived her of the private space which she had created for the interview. I realised that the interview is not going to be an easy one, as the interviewee could not immediately quit from the interview because, doing so would have generated suspicions that could have been harmful in her family setting. This judgment was not without basis, as reflected upon in my interview notes taken:

I could hear a male voice calling the interviewee’s name while knocking, and some steps walking towards the interviewee when the interviewee started talking again after a 10-second pause. The interviewee re-started the conversation with a completely different topic by saying that "I will be in the final year now once the exams are conducted". I told the interviewee, ‘You can opt out of the interview any time, even now’. (interview notes, 9 Dec. 2020)

My intention was to make her feel comfortable and prevent her from the guilt of leaving the interview if her family’s pressure demands. This active, culturally sensitive listening proved to be a key to confidence-building as ten minutes after the interview ended, the interviewee texted me, confirming the date for a follow-up interview, which had not been requested during our brief conversation.

Qualitative researchers have long been drawing attention to the benefits of effective listening for enhancing the quality of telephonic interviews. Farooq et al. (2017: 307) suggests that interviewers need to
listen to and pick up on changes in verbal cues such as 'pauses, hurried answers, tones, etc. And indicate if interviewees are confused, hesitating or experiencing frustration' (Farooq et al. 2017: 307). Auditory background cues are important along with verbal cues provided by interviewees in order to consider the context in which interviews—data collection—take place (Holt 2010). To understand this context for contextualising verbal and auditory cues, culturally sensitive listening and attention are important for deciphering verbal cues provided directly by the interviewee. Obdenakker (2006) highlights that invisibility of the interviewee to the interviewer juxtaposed with her visibility to the people surrounding her might reduce the possibilities the creation of a good interview ambience (and establishing rapport). Paying attention to such cues thus offers a potential to improve the interview ambience and subsequent rapport within the given setting and timing or by rescheduling the interview and thus allow for modifying/negotiating the settings in place (and thus the more than often restricted agency of female interviewees).

In the context of the case study site, the Swat Valley, I found that voice-only interviews enjoyed a clear advantage over face-to-face interview in terms of re-arranging the interview between a male researcher and a female interviewee given the strictly gender-segregated Pakhtun culture. Rescheduling a telephonic interview meant that female interviewees did not have to negotiate their own mobility to the interview site (as well as related timing issues/time constraints), as interviews with male researchers at women’s homes are considered culturally inappropriate/non-feasible. At the same time, this facilitated positive responses to a rescheduling of interviews in case the home setting for voice-only interview proved to be problematic. Transposing these concerns into research practice, great attention is warranted for the way cultural norms shape or sensor respondent’s answers, particularly in an artificial conversation setting (Au 2019: 60). The problems in this connection range from the positionality of researcher as a role, to the artificiality of the interview, to sudden shifts in how to word questions, all of which must be tailored to circumvent, rather than trigger, awkwardness, distress, and fear in the cultural norms adhered to by the respondents (ibid.: 61). While doing so, neglecting the role of cultural norms in reflecting on interview practices leave blind spots in assessing the reliability of data.

Culturally sensitive listening relaxes the interview dynamics both in the case of cross-cultural when the interviewer is an outsider (Sands et al. 2007: 364-6), an insider, or has a complexed multi-layered positionality like I do (regionally and culturally I am an insider, but I am an outsider as an employee of a UK-based university and a male researcher
interviewing women). The purpose here is not outlining a guide for culturally sensitive listening, but to argue that online interviews, conducted by researchers who are outsiders, insiders, or indigenous outsiders may improve by being cognisant of culturally sensitive listening. Culturally sensitive listening entails elements of 'reflective listening' (Au 2017: 64), effective listening, and culturally sensitive communications (Brooks et al. 2019). It is focused more on during-the-interview attention to interviewee’s change of tone, pauses, shifts in subject, not opening up and being or becoming unresponsive to certain subjects as a consequence of change in the interviewee’s physical space. In voice-only interviews, transformations in interviewee’s physical settings can be detected by paying attention to sounds of all kinds in the interviewee’s background, for instance, the opening of a door or voices in the background. At this stage, my data is unable to suggest a patterned relation between different kinds of sounds in the interviewee’s background and their effect over interviewee. However, I found that these sounds have implications for during-interview dynamics and analysing data generated from those interviews. Moreover, these sounds and their effects on the interview require interpretation within the cultural context of the interviewee.

**Voice-only interviews offer interviewees a greater control over interview settings**

The power of online communication technology to connect the domestic and the public sphere has allowed women to not only speak to male researchers, it also has empowered the women to speak to the male researchers on their own terms. Empowerment of women to speak on their own terms does not imply relinquishment of family’s influence over women’s use of mobile phone (Ibtasam et al. 2020). Instead, it denotes greater ability to choose interview time and space by the interviewee based on her sensibilities of her personal surroundings. After consenting to be interviewed, decisions related to when, how, and how long the interview would be, were exercised by the participants. In some instances, the interviewees bracketed the subjects which they were unwilling to talk over. However, these choices were not uniform; rather they reflected individual sensibilities of her surroundings and the subjects she deems worthy of conversation within these. Our topics of conversation included militancy and violence, cultural production, experiences of mobility in the marketplaces during and after the militancy in Swat, and memories and emotions surrounding these aspects of life during and after the conflict. Some made their consent conditional to the exclusion of subjects related to terrorism, while others wanted their stories amidst Taliban violence to be heard. Others avoided some topics
during the conversations such as feelings and emotions, and some others remained silent about any limitations either before or during the interview.

Telephone interviews also allowed women to select interview setting of their choice, hence they chose place and time of the interview in which they could talk freely. For instance, working women chose office hours so that they 'can talk freely without someone in the house hearing them' (interview notes, 3 Dec. 2020). In the case of non-working women where home was the only venue available to women interviewees, they created a favourable interview setting for themselves by suggesting appropriate time. For instance, a married university graduate, age 29, told one of my female gatekeepers,

Ask him [the author] to call me tomorrow any time after 10 a.m. and before 12p.m., because that is the time when our men are certainly out of the house. You [the female gatekeeper] know, when our men are around, we cannot talk to men over the phone. If they are not around, I will be comfortable in answering his questions. (reported by a female gatekeeper, 22 Sept. 2020)

Evidently, interviewees tried to create a private space for themselves by some of them locking the rooms while talking to the interviewer. This was noted earlier in section four in the case of the 23 years old university student. Although the knock on her door by a male family member deprived her of that space which led to a follow-up interview.

It is important to note that in face-to-face interviews between our female RA and women interviewees, ability to create a private interview space was limited compared to telephone interviews. In face-to-face interviews, the woman interviewee was always accompanied by her mother, sister, or another female relative, and such presence affected the responses offered by the interviewee to the interviewer’s questions. In her reflective notes on a face-to-face interview with a 25-year-old university graduate, Aneela, one of the research assistants, noted:

She [interviewee] was managing how to respond and took a deep breath, she looked at me [Aneela], then toward her [interviewee’s] mother and said (...). After completing her response, she looked toward her mother (for approval of her statement) and her mother also agreed with her point by knotting her head. (Aneela’s interview notes, 15 Dec. 2020)

Undoubtedly, such visual social cues are not available in voice-only interviews over the telephone or WhatsApp, nevertheless, voice-only interview give interviewees the choice and power to avoid such influences over their responses by managing the interview setting according to their own convenience.
Voice-only, synchronous interviewing not only enabled my women interviewees to decide their interview setting, but interviewees were also disclosing more information in online as opposed to face-to-face interviews. As noted in the methods section, I conducted most of the voice-only interviews alone with women interviewees on the thematic focus of our project on women’s markets in Swat. Two interviews were co-conducted with Aneela. One of these online (co-conducted) interviews was a follow-up to a face-to-face interview conducted by Aneela. The interviewee was providing more detailed responses than the face-to-face interview. As a way to formally close the online interview, I thanked the interviewee and apologised for taking more time than what was agreed in the follow-up request. The interviewee responded, 'I could have shared more stories with you [the author], but I thought that you will get tired by listening to my stories.'

Not all the online interviews were so revealing, for instance, in one interview, the interviewee was giving very short responses to all the questions, however, it was noted that her responses were particularly shorter when it came to questions about personal stories, but when she was asked to comment on collective experiences of violence, her responses were relatively longer (Interview notes, 10 Dec. 2020). Thus, voice-only synchronous interviews do not necessarily generate less content as Jonson et al. (2019 among others) suggests, the nature of questions such as personal vs. public, particular vs. General, and the influence of interviewee’s surroundings are greater influences on the length of responses than the mode of interviewing.

In fact, invisibility afforded by voice-only interviewing to the interviewees can elicit detail responses than face-to-face interview between a male interviewer and a female interviewee. In the cultural contexts such as Swat valley, where face-to-face meetings between men and women are against the social norms, women feel relatively comfortable in a voice-only interview with male researchers than a face-to-face, even if they have the opportunity to the latter. As stated in the methods section, after the seventh voice-only interview, I have started asking the interviewees to reflect upon if the voice-only interview was better for her, or face-to-face interview would have been easier? In the eighth interview, where Aneela and I were both present, the interviewee stated 'it [voice-only interview] is a lot easier over the telephone.' Aneela prompted, 'What if I would have come alone to conduct face-to-face interview?' The interviewee responded, 'For you [Aneela] and me [interviewee] it would have been easy, however, it [this particular interview] is easier through telephone.' Absence of additional concerns of women interviewees associated with face-to-face interviews such as 'covering my [interviewee’s] face appropriately,' not worrying about the
logistics of 'going to and coming back from the interview venue,' 'asking permission from my [interviewee’s] family for a face-to-face interview,' made voice-interview setting comfortable for them to articulate their responses.

**Advantages and caveats for interviewers using WhatsApp**

Instead of using landlines or cellular connections for telephone interviews, all the interviews were conducted over WhatsApp which has clear advantages over telephone calls. The advantages of using WhatsApp go beyond the preference of our women interviewees for an audio call, and the inexpensive nature of calls made over WhatsApp. In the context of Swat Valley, where our fieldwork is underway, WhatsApp is found to be accessible to a diverse range of population because of its user-friendly nature and its widespread usage in the Swati context. Undoubtedly, advances in technology shape the ways we do research, however, using online technologies for conducting qualitative interviews begs the question of 'how technology fits in the lives of the potential participants' (Sturges & Hanrahan 2004: 116; James & Busher 2006: 417).

This question should be at the heart of any online research in South Asia, which has the largest gender gap, 26 per cent, with respect to mobile phone ownership. In this connection, Pakistan is at the bottom among all the South Asian countries. According to the recent GCMA gender gap report, 44 per cent of Pakistani women have mobile phones compared to 80 per cent of Pakistani men (percentage of the adult population) (Ibtasam et al. 2020). The use of mobile phones among Pakistani women is on the rise, both in the urban and some rural settings for professional, educational, and business purposes (Aksar et al. 2020: 104; Sabit et al. 2020). Cultural barriers, including the belief that the internet is unsafe or that the use may expose women to outsiders continues to create hurdles in women’s use of technology. Moreover, affordability, lack of access to devices and connectivity and domestic responsibilities at home that leave women with no time of their own are also critical hurdles for women (Rehman et al. 2021: 6).

Some of these contextual specificities, such as technological competence of participants to communication technologies are not exclusive to women in the Swat Valley. This was evident in our fourth interview with a local male poet aged 60, who is a retired schoolteacher, a founder and active member of a local literary association. We (I and my colleague James) found out that the interviewee was not able to read the information sheet and consent form because of the unavailability of PDF reader on his phone. At the end of our hour-long interview, I said to the interviewee, 'If you have any questions or wanted to opt out of the
research, you can inform us as written in the consent form.' The interviewee informed the interviewers about the inability of his phone to read PDF. I asked, if emailing the documents would be helpful, and the response was in affirmative. However, the interviewee put the interviewers on hold and went into another room of his house and asked his wife about his son (name omitted), who was out with his friends. Exasperated about the call on hold, the interviewee asks his wife, 'Hey, do I have an email address?' (Interview, 9 Aug. 2020). This encounter with a male interviewee served as a warning to take the question of technical competence equally seriously as with the socially mediated access of women to communication technologies.

Sensitised by this encounter, the author of this paper sent a translated (in Pashto) version of the participant information sheet and consent forms in a voice note via WhatsApp to all of the women interviewees. In addition to the researcher’s introduction by the gatekeepers to the potential participants, this voice note served as an important introductory background to the researchers and the research they are conducting. Moreover, in some instances, the voice messaging function was used to ask follow-up questions. These voice messages are asynchronous of both time and place, but they have a clear advantage over written communication via email or other written methods as they can be used in the respondent’s language with lesser threat of miscommunicating ideas (James & Busher 2006; Obdenakker 2006). These voice messages also overcome, to some extent, the problems of bad connection as they are recorded clearly on both the sides (interviewer and interviewee) and delivered when the recipient is available online. Through this voice messaging function, the interviewer and interviewee stay connected in an asynchronous space which is particularly important for trust building and long-term relation between the researcher and research participants. Telephones do not offer such space. Although other mediums of communication such as skype and teams do, but they are not as widely prevalent as WhatsApp in the cultural geographical context where our research is underway.

Nevertheless, using voice messages for follow-up questions also comes with a caveat of compliance with ethical procedures. The researcher gets (by default) a recorded voice of the interviewee which stays in the WhatsApp chat until the interviewer deletes it. The interviewer should clearly inform the interviewees about how these messages will be treated. In my case, I deleted all the voice messages after listening to them and transcribing some of them verbatim.

Transcribing voice messages verbatim was a deviation from our initial stance (outlined in the methods section) on recording vs. note-taking.
Not all notes were transcribed verbatim, but the most important ones for analytical purposes. Unlike voice-only interview calls, where the auditory cues provided contextual richness in the absence of the visual cues, most of the voice notes (32 out of 35 at the time of writing) had a pin-drop silence in the background. Certain ethical challenges emerging from conversing with interviewees through voice-notes for any follow-up questions were considered prior to employing them as a supplementary component of inquiry. First, voice-notes were used as a method of last-resort. Second, in case of paramount necessity for using voice-notes, the interviewee was reminded in advance that these asynchronous voice-notes are automatically recorded against the wishes of the interviewer and interviewee. The interviewees were asked to use "delete for everyone" function after four minutes of the note being sent. "Delete for everyone" removes the content both for the sender and the recipient, however, this option is available to the sender for a limited time. Nevertheless, the interviewer deleted every voice-note of the interviewee, to avoid confusing the interviewee’s loss of opportunity to "delete for everyone" with her intentional non-deletion.

In addition to informing interviewees how these messages will be treated, researchers are warned to pay special attention to data protection amid emerging debates on increasing uncertainty around WhatsApp’s new privacy policy (We can’t see your private messages, 2021). This is especially important for researches on sensitive topics and vulnerable populations.

Conclusion

For arguments surrounding voice-only interviewing methods, to borrow Bisoka’s words, 'Covid-19 is not an event, it is a reminder of the actuality of such debates.' Covid-19 is not an event, in so far as it does not truly deviate from the normal state of affairs (ibid.: 3). Researchers have long been calling for the need to enhance our understanding of what differences are made to our research if we choose, for whatever reasons, to conduct voice-only interviews and consider the implications of this mode-selection for our research practice (Irvine et al. 2011: 205). The increasing use of online communication technologies in the wake of Covid-19 makes it hard to believe that voice-only online interviewing will fall out of use (Sy et al. 2020), instead, they are likely to become more viable and practical options. This is particularly evident in our case of the gendered research geography of the Khyber Pakhtunkhwa province of Pakistan where voice-only interviews may allow researchers to cross gender boundaries more frequently for interviewing participants. Insider’s knowledge of the context especially about the
characteristics of potential participants is a key to successful online interviews (James & Busher 2006). Unlike James and Busher (ibid.), I did not personally know each potential participant, and neither I nor my project’s team members were strangers as they (accept James and Sarah) were the part of that cultural milieu, and had not recruited participants without the trusted gatekeepers who had the knowledge about each of the research participants recruited. Therefore, it is important that the researchers must recognise what they are doing when they do it, and what it means to take data at its face value (Sandelowski 2002).

This paper has demonstrated the utility of voice-only interviews in accessing hard-to-reach research participants in the strictly gender-segregated cultural context of Swat Valley. Voice-only synchronous interviews are more appropriate as a methodological necessity than a choice if the researchers are planning to recruit participants from opposite sex in comparable cultural contexts. In the case of women interviewees and male interviewers, voice-only interviews are more helpful in recruiting participants than online modes of video interviewing technologies. For this reason, I emphasise the significance of culturally sensitive listening (taking cues from the sounds in the background of the interviewee) to contextualise her verbal cues. This is to say, effective listening alone is not enough, effective listening is enriched by culturally sensitive listening in online interviews where visual social cues are absent. A methodological caveat is in order here. Methodological reflections of this paper do not extend to non-educated women in the cultural geography of Swat Valley, therefore, it remains to be seen how those women can be recruited in online interviews, and what lessons can we learn from employing culturally sensitive listening in online interviews with non-educated women.

This is where Covid-19 can be viewed as an event for using remote interviewing methods by qualitative researchers and appropriateness of these methods for facilitating outreach across gender boundaries. My reflections are not aimed at providing a methodological guide, but a pointer towards a promising methodological direction for future research. Pandemic or not, the question for gender positionality of researchers and its implications for sampling frames begs serious and hither to unanswered questions. For instance, are pre-conceived notions of cultural norms underpinning justification of ex-post sampling closure or careful vetting of alternatives is reflected in the research design?
## Appendix 1: Sample Characteristics

<table>
<thead>
<tr>
<th>Interviewee Education</th>
<th>Age</th>
<th>Married?</th>
<th>Village or city</th>
<th>Face-to-face or voice-only</th>
<th>Interviewer</th>
<th>Place of interview</th>
<th>Who facilitated the contact?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double Masters</td>
<td>30 Y Village</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's Home</td>
<td>Author's acquaintance and interviewee's colleague (male)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>50 Y Village</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's Home</td>
<td>Interviewee's and author's mutual acquaintance (male)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>55 Y Town/city</td>
<td>Face-to-face</td>
<td>Research assistant</td>
<td>Interviewee's home</td>
<td>male interviewee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>38 N Village</td>
<td>Face-to-face</td>
<td>Research assistant</td>
<td>Interviewee's office</td>
<td>Research assistant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>40 Y City</td>
<td>Face-to-face</td>
<td>Research assistant</td>
<td>Interviewee's office</td>
<td>Research assistant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No education</td>
<td>45 Y Village</td>
<td>Face-to-face</td>
<td>Research Assistant</td>
<td>Interviewee’s home</td>
<td>female interviewee</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate</td>
<td>21 N Town</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's home</td>
<td>Author’s friend and interviewee’s acquaintance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masters</td>
<td>25 N Village</td>
<td>Face-to-face</td>
<td>Research Assistant</td>
<td>Interviewee’s home</td>
<td>Author’s acquaintance and interviewee’s female cousin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>Degree</td>
<td>Age</td>
<td>Gender</td>
<td>Location</td>
<td>Methodology</td>
<td>Interviewee's Home</td>
<td>Source</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>-----</td>
<td>--------</td>
<td>----------</td>
<td>-------------</td>
<td>-------------------</td>
<td>--------</td>
</tr>
<tr>
<td>9</td>
<td>Masters</td>
<td>26</td>
<td>N Village</td>
<td>Face-to-Face</td>
<td>Research Assistant</td>
<td>Interviewee's home</td>
<td>Author's acquaintance and interviewee's brother</td>
</tr>
<tr>
<td>10</td>
<td>Primary</td>
<td>50</td>
<td>Y Village</td>
<td>Face-to-face</td>
<td>Research Assistant</td>
<td>Interviewee's home</td>
<td>Author's acquaintance</td>
</tr>
<tr>
<td>11</td>
<td>Masters</td>
<td>45</td>
<td>N Village</td>
<td>Voice-only</td>
<td>Author and research assistant</td>
<td>Interviewee's home</td>
<td>Research assistant</td>
</tr>
<tr>
<td>12</td>
<td>Undergraduate student</td>
<td>19</td>
<td>N village</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's home</td>
<td>Author's acquaintance and Interviewee's teacher (male)</td>
</tr>
<tr>
<td>13</td>
<td>Masters</td>
<td>38</td>
<td>N Village</td>
<td>Voice-only</td>
<td>Author and Research assistant</td>
<td>Interviewee's home</td>
<td>Research assistant</td>
</tr>
<tr>
<td>14</td>
<td>Masters</td>
<td>45</td>
<td>-- village</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's home</td>
<td>Author's acquaintance and interviewee's friend (female)</td>
</tr>
<tr>
<td>15</td>
<td>Masters</td>
<td>29</td>
<td>Y Village</td>
<td>Voice-only</td>
<td>Author</td>
<td>Interviewee's home</td>
<td>Author's acquaintance and interviewee's friend (female)</td>
</tr>
</tbody>
</table>

**Endnotes**

1 Disembodied here implies the absence of physical exchanges that occur in a reciprocal manner between the researcher and the participant, and that has implications at the "physical level." Thus, the body (of the researcher and the participant) as an instrument of research and an anchor of
reflexivity is absent in an online voice call. For interviews as embodied communication, see Burns 2003. For a feminist critique of relying on disembodied online research see, Morrow, Hawkins & Kern 2015.

2 Throughout this paper, I use singular and collective personal pronouns to reflect upon actions related to this project. Wherever I use collective personal pronoun, it refers to actions decided by the project’s team collectively.

3 This term is used as a shorthand to refer to cultural geographies where social norms related to interaction across gender boundaries constrain research designs to recruit participants from the opposite gender to that of the researcher.


5 Our research assistants were sent to the field in November and the first week of December before the second wave of Covid-19 started hitting Pakistan. Since 5 December, 2020, to the time of writing (May 2021) our research assistants have been working from home.

6 This procedure of providing translated information sheet and consent forms, and withdrawal of participants is documented on p. 12 of our ethics application approved by the Kings College London Institutional Review Board.

7 I thank Izhar Khan, a local cab driver and my friend for transporting the physical copies where required.


Bibliography


Seitz, Sally. 2016. Pixilated partnerships, overcoming obstacles in qualitative interviews via Skype: a research note. *Qualitative Research*, 16 (2), pp. 229-35.


