EFFECTS OF DIFFERENT TYPES OF FACE COVERINGS ON SPEECH ACOUSTICS AND INTELLIGIBILITY

CARMEN LLAMAS, PHILIP HARRISON, DAMIEN DONNELLY AND DOMINIC WATT

Abstract

This paper reports the results of two experiments investigating the effects on speech acoustics and intelligibility of a number of different types of forensically-relevant fabric mouth and face coverings, including the *niqāb* (full-face Muslim veil), balaclava, and surgical mask. For the perceptual (intelligibility) experiment, subjects were presented with two types of speech stimuli, 'bimodal' and 'unimodal', and asked to write down what they heard. Four facial guises were used (*niqāb*, balaclava, surgical mask, no covering). In the bimodal condition (video + audio), subjects saw and heard video recordings of actors reading target words embedded in a standardised carrier sentence. In the unimodal (audio only) condition, subjects heard just the soundtrack of the same video recordings, i.e., no visual image was present. It was found in the perceptual test that the subjects could in all four guise conditions correctly identify target words with a high degree of reliability, and that a small number of confusion types accounted for the majority of the errors. For the second (acoustic) experiment, the objective was to assess the sound transmission loss characteristics of the fabrics from which these and other face coverings are composed. This experiment showed that transmission loss was negligible for all but one of the fabrics, suggesting that speech intelligibility problems created when mouth and face coverings are worn by speakers must derive principally from the reduction in visual information available to listeners and/or from the auditory consequences of interference with speech articulation caused by the face coverings, rather than from transmission loss of the fabrics themselves.

1. Introduction

Our interest in this topic stems originally from the controversy that arose from comments made in an article in the *Lancashire Telegraph* newspaper (6th October 2006) by Jack Straw MP, the former UK Home Secretary (1997-2001), the current Secretary of State for Justice and Lord Chancellor, and the then (as now) Member of Parliament for Blackburn. A high proportion of Straw's Blackburn constituents are Muslims of south Asian origin, and many women from that community habitually veil themselves in garments such as the *niqāb* (full-face veil, with a slit for the wearer's eyes) or *burqa*, which covers the entire head and body, and which often has a fabric mesh obscuring the wearer's eyes as well. Straw remarks in the article that he found it difficult to hold conversations with female constituents who came to his surgeries (consultation hours) dressed in garments of this type.

> I felt uncomfortable about talking to someone "face-to-face" who I could not see. [...] I think... that the conversation would be of greater value if the lady took the covering from her face. Indeed, the value of a meeting, as opposed to a letter or phone call, is so that you can - almost literally - see what the other person means, and not just hear what they say. So many of the judgments we all make about other people come from seeing their faces.
>
> (*Lancashire Telegraph*, 6th October 2006)

Straw does not appear to be claiming that he experiences difficulty hearing what his veiled constituents are saying; rather, he is drawing attention to the contribution made by visual information when processing conversational speech, and to the fact that an absence of visual cues may hinder recovery of the talker's intended message. In making these observations Straw may have underestimated the acute sensitivity of certain members of the Muslim community to perceived attack, and the uproar which followed the publication of his article took several weeks to subside.

The inflammatory situation was not helped by the fact that less than two weeks after Straw's article was published Aishah Azmi, a language support worker employed by a school in Dewsbury, West Yorkshire, lost the discrimination/harassment case she had instigated in response to her having been suspended from her post. The school and the local authority Kirklees Council had argued that Azmi's refusal to take off her *niqāb* in the classroom if a male colleague was present meant that pupils in her care would be unable to understand what she was saying. It is unclear whether this argument rested on the claimed unintelligibility of Azmi's speech caused by disruption to visual cues, or on degradation of the auditory signal brought about by Azmi's mouth and nose being covered by the veil.

The issue of the *niqāb*'s potential to impair communication again made media headlines in the wake of an incident during an immigration case being heard on November 6th 2006 in Stoke-on-Trent. Shabnam Mughal, a legal executive from Coventry, was asked by the judge, George Glossop, to remove her veil during the court proceedings, so as to "assist communication" and because, in Glossop's words, "I cannot hear you as well as I would like" (*Daily Telegraph*, 8th November 2006). Mughal twice refused to comply, and the case was adjourned so that Glossop could seek advice from Mr. Justice Hodge, President of the UK Asylum and Immigration Tribunal. Hodge's guidelines, issued on 9th November, state the following:

> Immigration judges must exercise discretion on a case-by-case basis where a representative wishes to wear a veil. The representative in the recent case has appeared veiled previously at AIT [Asylum and Immigration Tribunal] hearings without difficulties. It is important to be sensitive in such cases. The presumption is that if a representative before an AIT tribunal wishes to wear a veil, has the agreement of his or her client and can be heard reasonably clearly by all parties to the proceedings, then the representative should be allowed to do so.
>
> If a judge or other party to the proceedings is unable to hear the representative clearly then the interests of justice are not served, and other arrangements will need to be made. Such arrangements will vary from case to case, subject to judicial discretion and the interests of all parties.
>
> (Judicial Communications Office, News Release 29/06)

In July 2007 further guidelines were circulated by the Equal Treatment Advisory Board Committee of the Judicial Studies Board. These guidelines, prepared by the committee Chair Mrs. Justice Cox, advised that each case should be considered individually, that judges should assume female Muslim lawyers are entitled to wear the veil, and that it will not automatically create any difficulties. As stated in the *Equal Treatment Bench Book* (2007: 3-18/1),

> it is for the judge, in any set of circumstances, to consider what difference, if any, would be made to [the interests of justice] by the niqab being worn.

Furthermore,

> In reality, in the absence of any question relating to identification, there are few instances where an advocate or representative appearing in a niqab would be likely to present any real issue. Such concerns would be likely to centre round the fact that the woman could not be heard, rather than seen.
>
> So long as the advocate can be heard reasonably clearly it is unlikely that the interests of justice will be impeded. Just as in any case where a judge might have difficulty in hearing any party, witness or advocate, sensitively enquiring whether they can speak any louder or providing other means of amplification should suffice and such measures should be considered with the advocate before asking her to remove her veil.
>
> (Equal Treatment Bench Book 2007: 3-18/5)

The debate is periodically revived in the mainstream media (e.g., in reaction to the speech on women's human rights in the 21st century made at Chatham House by Cherie Booth QC in October 2007) but at no point since the publication of the Straw article, or indeed before that, has any objective evidence been put forward either to support or refute the conjecture that the spoken language of *niqāb* wearers is in fact more difficult to hear or understand than that of talkers with uncovered faces.

To gather evidence of this sort, then, is a key purpose of the study reported here. We acknowledge that there is ample scope for study of the effects of garments such as the *niqāb* and *burqa* on interpersonal communication (e.g. how turn-taking in conversational interaction is handled, and the like), and would strongly encourage other researchers to begin systematic investigation of the sociolinguistic and sociopsychological implications of the wearing of face coverings of this type. The latter could be conducted via interviews or questionnaires administered to habitual wearers, to those living in communities in which wearers are regularly encountered, and to those who have little experience of interacting with wearers. For the initial stages of our own research, however, we decided to follow a more formal experimental approach according to which the factors we hypothesised to be relevant could be relatively tightly controlled.

We describe below two experiments carried out to assess the extent to which speech intelligibility and speech acoustics are affected when fabrics of various types intervene between a sound source (talker's mouth/nose or a loudspeaker) and the ear of the listener. These fabrics are of three types: (i) fabrics made into garments designed to be worn as headgear or as a mask to cover the talker's mouth and nose; (ii) other fabrics which in certain circumstances might be used to cover the talker's mouth or entire face, or perhaps to cover the mouthpiece of a telephone, for example during an armed robbery or an anonymous phone call; and (iii) so-called acoustically transparent material of the sort used to cover loudspeakers. As should by now by clear, the choice of fabrics tested was motivated chiefly by forensic phonetic considerations, as we wished to address questions relating to the intelligibility of speech produced by talkers whose mouths or entire faces are obscured by fabric coverings of the sort that might be involved in situations having a legal aspect. Aside from the real cases discussed earlier in this section, one can readily imagine other scenarios in which a garment worn on the talker's face might be the focus of complaints based on claims about impaired speech intelligibility (e.g. in medical or military contexts, among emergency services workers, in manufacturing plants such as foundries or electronics factories where face masks of different types are worn, and so forth).

The research questions we chose to attempt to answer in the present study are discussed in the following section.

2. Research questions

The specific research questions that arise from the issues discussed in the preceding section are of two sorts: those under (1) relate to 'viewer-listener' exposed to speech produced by a person wearing a *niqāb*, and question (2) relates to the properties of the fabric itself.

(1) a. Is speech intelligibility adversely affected if the talker is wearing a *niqāb*?
(1) b. To what extent is any such effect the result of disruption to/absence of visual cues?
(1) c. To what extent is any such effect the result of disruption to/absence of auditory cues?
(2) With respect to the point raised in (1c), how does the *niqāb* compare acoustically with other forensically-relevant fabrics used in mouth and face coverings and with a control condition in which no fabric is present?

In section 3 we address questions 1(a-c) by discussing the background to the perceptual (intelligibility) experiment, the methods we employed to investigate it, and lastly the results of that experiment. In section 4 we describe previous research on sound absorption (transmission loss) characteristics of fabric types including that used for the facial veil part of the *niqāb* worn by the talkers in the video recording made for the perception experiment, and give details of the methods and results for the acoustic experiment run on the fabric samples. Finally, in section 5 we suggest directions for future research in this area.

3. Perception Experiment

3.1. The role of vision in speech perception

There exists a very sizeable body of research literature on the contribution of visual information to speech perception (see e.g. Summerfield 1987; Massaro 1987; Green 1998; Schwartz, Robert-Ribes, and Escudier 1998; Campbell 2008). Speech perception has in fact been described as the prototypical form of multimodal perception (Massaro 2001: 14872), and given the acoustic-auditory focus of much of the speech perception research published to date it would be easy to underestimate the centrality of the role played by visual cues in face-to-face interaction. A lack of space prevents the inclusion here of anything but a review of rather a selective subset of the relevant literature, however. For our purposes the most relevant sources are those which deal with bimodal (auditory and visual) speech perception in visually degraded conditions, specifically those in which part or most of the visual image accompanying an auditory stimulus is obscured or removed.

Early experimental work by Cotton (1935) and Sumby and Pollack (1954) demonstrated that being able to see a speaker's mouth movements greatly enhances intelligibility, especially where speech is presented in noise. This general finding has been replicated in numerous subsequent studies (see e.g. Remez 2005; Campbell 2008). Visual cues to speech articulation may even be sufficient to override auditory cues, as demonstrated by research on the McGurk Effect and related phenomena (McGurk and McDonald 1976; Jones and Callan 2003; Brancazio and Miller 2005), while non-linguistic visual information has been found to influence auditory judgments in unexpected and subtle ways (e.g. Strand and Johnson 1996; Johnson, Strand, and d'Imperio 1999; Glidden and Assmann 2004; Hay, Warren, and Drager 2006).

Research specifically on the contributions made to speech intelligibility by movements of the mouth and the face more generally has shown that viewer-listeners can recover relevant information from the talker's face with high levels of success even when the visual image is manipulated or degraded (e.g. Berger, Garner, and Sudman 1971). Massaro (2001), for instance, comments that

> ... speech reading, or the ability to obtain speech information from the face, is not compromised by oblique views, partial obstruction or visual distance. Humans are fairly good at speech reading even if they are not looking directly at the talker's lips. Furthermore, accuracy is not dramatically reduced when the facial image is blurred (because of poor vision, for example), when the face is viewed from above, below, or in profile, or when there is a large distance between the talker and the viewer.
>
> Massaro (2001: 14873)

As one might anticipate, the lower part of the face appears to contribute relatively more useful information during speech production than does the upper part, at least for tasks involving word recognition. In the domain of intonation, however, Lansing and McConkie (1999:536) suggest that information is distributed more broadly, such that viewer-listeners may be able to recover useful signals relating to the prosodic structure of utterances on the basis of movements of the eyebrows and head. Using eye-tracking techniques to monitor which parts of the talker's face viewer-listeners fixated on during exposure to bimodal stimuli, Munhall and Vatikiotis-Bateson (1998) sought to identify which regions of the face are the most informative during speech perception. Their subjects viewed recorded monologues with varying levels of noise superimposed on the soundtrack. The size of the facial image was incrementally varied (see also Jordan and Sergeant 1998, who employed a similar technique). Munhall and Vatikiotis-Bateson found that the image size did not affect the fixation pattern, but that the noise level did: subjects looked more frequently at the talker's mouth as the signal-to-noise ratio of the soundtrack was reduced. The mouth region of the talker's face was fixated on only around half of the time even in the noisiest condition, however, which, according to Munhall and Vatikiotis-Bateson, would indicate that subjects were attempting to recover relevant information from other parts of the talker's face (in particular the eyes), or possibly that they were taking advantage of the better motion acuity in the peripheral visual field to detect small movements of the lower part of the face by fixating on the eyes. They suggest also that the talker's eyes might provide a useful perceptual 'anchor point' to help counteract spatial tracking problems caused by simultaneous movements of the talker's jaw and head.

Speech perception studies in which parts of the talker's face are obscured are relatively rare. Erber (1979) placed rough-surfaced Plexiglass between viewer-listeners/(deaf) lipreaders and the talker, and increased the distance between the Plexiglass and the talker in a way that progressively enlarged the amount of distortion in the visual image. A ceiling effect whereby performance in word-recognition tasks reached a plateau showed that subjects with normal hearing came to rely exclusively on auditory information when optical distortion reached a certain threshold. The results for the hearing-impaired subjects (all children) indicated that lipreading became impossible relatively quickly as the Plexiglass sheet was placed further from the talker, though the presence of bilabial articulations in some of the target forms predicted relatively high levels of successful identification as the level of optical distortion was increased.

More recently, Rosenblum and Saldaña (1998) have employed a 'point light' technique, whereby small fluorescent markers were placed on a talker's lips, teeth, tongue, and face in three different configurations. With black make-up applied in relevant areas inside the mouth and on the face, and using special lighting conditions, only the markers (or ‘points’) are visible in the video image. Rosenblum and Saldaña found that all three configurations of points significantly assisted adults with normal hearing to correctly identify sentences played in white noise, though there appeared to be little improvement in performance when points were placed on areas of the face other than the lips, teeth and tongue.

The findings discussed above confirm the common-sense prediction that the most important part of the face in terms the contribution made to speech intelligibility by visual cues is the mouth. Rather few studies have, however, looked at speech production in ‘realistic’ conditions, whereby the mouth and/or face are obscured by headgear or facemasks. Those studies which have been conducted focus on impairments to speech intelligibility caused by masks of the sort used by military and emergency services personnel, deep-sea divers, and so forth (see e.g. Bond, Moore, and Gable 1989; Bishop, Bahr, and Gelfer 1999), though these are often worn in conditions which would impair speech intelligibility otherwise (e.g. because of high-levels of background noise such as machinery noise or gunfire, through signals being relayed through radio or telephone links, through wearers breathing heliox, etc.).

Coniam’s (2005) study of Hong Kong students during the 2003 outbreak of Severe Acute Respiratory Syndrome (SARS) is the only study we are aware of that seeks to address research questions similar to our own. During the outbreak, large numbers of people in Hong Kong attempted to protect themselves from being infected by the SARS virus by wearing surgical masks while out in public. Coniam enlisted students to take part in the oral elements of a mock Hong Kong Certificate of Education Examination while wearing surgical masks, and found that their performance was not markedly impaired, except in their marks for pronunciation in the role play section. A number of Coniam’s subjects reported that they had adopted compensatory strategies to overcome the effects of the mask, such as ‘speaking more loudly or attemping to make greater eye contact or use more body language’ (Coniam 2005: 252). These attempts may account for the fact that raters asked to evaluate recordings of the mask-wearing subjects were equivocal as to whether wearing a mask had made any difference to the test-takers’ performance.

As mentioned in section 2, we have been unable in the work reported on in subsequent sections to make systematic observations of compensatory articulatory or gestural strategies parallel to those reported by Coniam’s subjects. In the following section, however, we describe the methods used for the perceptual experiment which was designed in order to assess the extent to which wearing a *niqāb*, a balaclava or a surgical mask impaired speech intelligibility.

3.2. Read materials

Forty monosyllabic English target words, all but one with a CVC syllable structure, were chosen to exemplify a range of consonants (/p t k s ʃ z f v h θ ð tʃ dʒ m n/) in onset (/p t k s ʃ z f v h θ ð m n/) and coda (/p t k s ʃ θ tʃ dʒ m n ŋ/ + /ts/) positions. Four vowels (/i ɪ a ɔ/) were exemplified. The choice of words was partly motivated by their potential confusability–minimal and near-minimal pairs were thought most suitable for our purposes (see Miller and Nicely 1955)–and partly by their relative frequency (with certain exceptions such as *hick*).

pat	tat	cat	pip	tip
kip	sip	ship	zip	seat
sheet	fin	thin	fought	thought
pick	tick	kick	Vic	mat
mats	mass	mash	match	math
Madge	hip	hat	hick	hot
heat	hit	him	kin	Kim
king	nit	mitt	vat	that

Table 1: 40 target words read in carrier sentence *Say X again* for video recording.

The target words were embedded in the carrier sentence *Say X again* (where *X* is the target word) and prepared for reading in the form of a Microsoft PowerPoint slideshow presentation. The presentation was set to run automatically, with 2 seconds being allotted to each slide and a 4-second pause being inserted after each block of 10 slides. The sentences were read aloud by two actors in each of four guises, i.e. 8 times altogether, with the stimulus order randomised for each reading (see below).

3.3. Video recording

Two Standard Scottish English speaking actors, a 23-year-old male and 25-year-old female who were at the time both students at the University of Aberdeen, read the target utterances from the PowerPoint slideshow running on a laptop computer. This was placed directly below the lens of a Sony PD170 professional digital camcorder so that the actors' gaze line was as close as possible to the camera. The camera, which had a Sony shotgun electret condenser microphone attached, was mounted on a tripod, the positions of the legs of which were marked on the floor. A backup copy of the sound recording on a second channel was obtained by attaching a lapel microphone to each actor while they read the materials. The actors sat approximately 1.5m from the camera in a chair, the position of which had also been marked on the floor in case it was inadvertently moved. The actors were instructed to look straight ahead without moving their head or eyes and to vary as little as possible in volume and intonation as they read the test utterances. Owing to a difference in height between the male actor and the female actor the camera was adjusted very slightly so as to keep the actors' head and shoulders in the same place in shot.

Illumination was from above and was held constant by using only artificial light in a windowless room. The room used was not soundproofed, but pains were taken to ensure that as little ambient and extraneous noise as possible would appear on the recording. Air conditioning and all electrical/electronic equipment other than that needed for the experiment were switched off. All staff in the building were asked to refrain from talking outside the room and from letting doors close noisily.

3.4. Speaker guises

The actors were recorded talking in four guises: *niqāb*, balaclava, surgical mask, and no face covering. The *niqāb* was made of a light woven black polyester fabric thin enough to be partially translucent to light. The balaclava was composed of a medium thickness knitted acrylic yarn, and was of the type that covered the entire face but for the eyes. The surgical mask was of the disposable pleated-paper variety with elasticated ear-loop ribbons attached. Both talkers wore precisely the same garments during their readings in the three face covering guises. Neither of them had worn any of garments of these types prior to the practice runs that preceded the video recording session itself, so to that extent it is possible that their speech articulation during the recording was less than fully natural. If the experiment were to be replicated it would be advantageous to use talkers more accustomed to wearing one or other of the garments. Finding individuals who habitually wear more than one of the garments would of course be very difficult, perhaps impossible, so the use of actors who had had brief experience of wearing all of the garments seemed a reasonable compromise.

3.5. Preparation of the stimulus materials

After recording, the video was edited using Apple *iMovie* by separating each of the 8 actor/guise combinations into individual files. These files represented the bimodal (audiovisual) condition. They were then duplicated and the video image stripped out so as to create a second set of stimuli (unimodal, i.e. the audio-only condition) for presentation to experimental subjects. A total of 640 test utterances (2 actors × 4 guises × 2 conditions × 40 target words) blocked by actor/guise combination were thus produced. These were arranged in random order and burned onto DVDs for administration via individual PCs and headphones. The order in which they were presented to subjects is shown in Table 2.

Talker	no covering	*niqāb*	surgical mask	balaclava
male	audiovisual (9)	audiovisual (11)	audiovisual (3)	audiovisual (13)
	audio (6)	audio (8)	audio (16)	audio (2)
female	audiovisual (1)	audiovisual (15)	audiovisual (7)	audiovisual (5)
	audio (12)	audio (4)	audio (14)	audio (10)

Table 2: Sixteen blocks of combinations of three variables: sex, guise and audiovisual/audio conditions. Numbers in parentheses refer to the presentation order of the blocks.

3.6. Participants: viewer-listeners

Thirteen viewer-listeners participated in the perceptual experiment (10 female, 3 male; age range 18-37). Three had first languages other than English (French, Greek, Luxembourgish). All had lived in Aberdeen for at least three years so were familiar with Scottish English. They were given answer sheets with spaces provided for recording test words as they were heard (see example in Appendix A). Subjects wore headphones connected to the PC sound card. They were given instructions to watch the computer screen and, ignoring the other words in the carrier sentence, to write down on their answer sheet the target words as they heard them. An additional questionnaire (Appendix B) was administered after the experiment. In this, subjects were asked to rank the four guises from easiest to most difficult to understand, to mention any other factors that they felt affected their performance, and to circle any of the target words with which they were unfamiliar.

3.7. Results of perceptual experiment

The total number of misperceptions in the perceptual experiment was surprisingly small. Furthermore, it was decided that certain types of errors should not be considered. Spelling errors were discounted (although pseudo-phonetic representations of the test words were deemed acceptable); words marked as unfamiliar to listeners were discounted; subsequent misperceptions of words that were misperceived in the control guise (audiovisual, no covering) by the same subject were discounted; and vowel misperceptions were also discounted, given that the focus of the experiment was consonantal (mis)perception. After eliminating examples of these error types, the number of misperceptions thought suitable for further analysis was very small. Of the 8,320 tokens heard in total (640 stimuli × 13 listeners), only 165 misperceptions were recorded (i.e. just under 2%). The most common misperceptions were as follows:

- confusion of stops with (mostly homorganic) fricatives, especially /t/~/θ/
- place of articulation of stops
- voicing of stops
- place of articulation of fricatives, especially /f/~/θ/
- place of articulation of nasals, especially /n/~/ŋ/

A full list of the misperceptions may be seen in Appendix C.

Misperceptions that were reported for the same target word by three or more subjects – we label these ‘consistent misperceptions’ – are shown in Table 3 below.

		male							female						
		audiovisual			audio				audiovisual			audio			
order of presentation →		3	11	13	6	16	8	2	7	15	5	12	14	4	10
test word	perceived	surgical mask	*niqāb*	balaclava	no covering	surgical mask	*niqāb*	balaclava	surgical mask	*niqāb*	balaclava	no covering	surgical mask	*niqāb*	balaclava
tip	*pip*									3				1	
pick	*big* *pig*										1 1				7
thought	*fought*								3	1			2	1	
mats	*mat*	2		1		1		5	1				1		
match	*Madge*									3				4	
hick	*hip*								3		2		3		4

Table 3: Frequency of perceptual errors among panel of 13 viewer-listeners (includes only ‘consistent’ misperceptions, i.e. target words misidentified by 3 or more subjects)

It is immediately clear from Table 3 that a considerably greater number of consistent misperceptions are reported by subjects in response to the female speaker than is the case for the male speaker. In fact, of the total of 165 reported misperceptions, 115 were in response to the female speaker’s utterances, as opposed to only 50 in response to the male’s utterances.

What is also apparent from Table 3 is that the *niqāb* does not yield consistent misperceptions in the male speech. Only 5 of the 50 misperceptions recorded in total for the male speaker resulted from speech produced in the *niqāb* guise (3 in the audiovisual condition and 2 in the audio-only condition). The balaclava guise results in the highest number of consistent misperceptions for the male speaker. For the female speaker, the *niqāb* guise produces the same number of consistent misperceptions as the surgical mask, though this is smaller than the number reported for the balaclava guise. The latter again results in the highest number of consistent misperceptions subjects reported. By comparison with subjects' responses to the male's *niqāb* guise speech, however, the *niqāb* adversely affect word-identification accuracy much more markedly for the female's speech (17 misperceptions in the audiovisual condition, 21 in the audio-only condition).

The greater number of consistent misperceptions produced by the balaclava guise than by the other coverings in both the male and the female speech corresponds also to the listeners' reactions to the guises in the short questionnaire subjects completed after the listening session. The majority of subjects ranked the balaclava guise as the most difficult to understand. The surgical mask was ranked 'easiest to understand' of the three coverings, while the *niqāb* fell between the two.

For both speakers, a higher number of consistent misperceptions were reported in the audio-only condition as compared with the audiovisual. In terms of all 165 misperceptions, the audio-only condition increases misperceptions in female's speech by nearly half (68 misperceptions in the audio-only condition, as compared with 47 in the audiovisual). As the audio signal heard by subjects is identical in both conditions, this suggests that visual information is important for perception even in the masked guises.

Overall, the results of the perceptual experiment revealed unexpectedly high accuracy in the subjects' responses which may, in part, have been helped by familiarisation with the task as the listening session progressed. However, the effects of fatigue brought about by performing this quite lengthy and monotonous task may have counteracted any familiarisation-related improvements in performance. The misperceptions reported in the experiment demonstrated that the female speaker's productions were much less reliably perceived by listeners than were the male speaker's. The relative unintelligibility of the female speaker's utterances in masked guises (especially in the audio-only condition) may perhaps result from a greater impedance of the female speaker's articulators by the face covering than was the case for male speaker. This may be particularly true of the balaclava, which touches the speaker's lips to a greater degree, and at somewhat higher pressure, than do the two other face coverings used. Also, it may be the case that of the three face coverings the balaclava is responsible for greater damping in the higher frequency bands occupied by greater amounts of linguistically-relevant information in female speech than is the case in adult male speech (see Section 4 below).

Additionally, the general qualities of the female speaker's speech may have been responsible for the differences between the subjects' perceptions of the female speaker's speech compared with the male speaker's. The female speaker generally spoke more quietly than the male speaker and her speech may also have been less carefully articulated. Although the speakers were instructed to vary as little as possible in their delivery of the test utterances, a certain amount of inter- and intra-speaker variability must be considered an inevitable consequence of the nature of the experiment. The results must therefore be interpreted with variability of these types in mind.

3.8. Acoustic explanations for misperceptions

While we had complete control over the signals played to the subjects who participated in the perception test, we have no evidence of what the subjects perceived other than the written forms they submitted on their answer sheets. We cannot know for certain, therefore, what the source of their misperceptions might have been. It seems reasonable nonetheless to speculate about what sorts of misperceptions might have been predicted when the acoustic qualities of certain consistently misidentified stimuli are examined. A very obvious example involves the relatively frequent (×7) misidentification of *match* as *Madge* in the female *niqāb* guise in both audio-visual and audio-only conditions (see Appendix C). The periodicity observed in the section of the acoustic waveform corresponding to the stop closure for /tʃ/ in *match* confirms the authors' auditory impression that the affricate is indeed voiced, albeit weakly and only during the stop closure. Voicing of intervocalic voiceless coda consonants is common among English speakers, and it is therefore hardly surprising that listeners confused these two words. We should in this case not attribute the confusion to the effect of the face covering, but rather simply to a normal assimilatory process of English.

The spectrograms in Figure 1 provide a better illustration of how the acoustic influence of wearing a covering over one's mouth/nose can result in consonant misidentification. Three subjects reported hearing *hip* rather than the target word *hick* in both the audio-visual and audio-only conditions of the female speaker's surgical mask guise, as did five subjects in the female's balaclava guise (2 in the audio-visual condition, 3 in the audio-only condition). It can be seen that the release burst for /k/ in the 'no cover' guise spans the frequency spectrum and is followed by a significant period of fricative noise. Nothing comparable can be seen in the spectrogram for the /k/ of *hick* in the surgical mask guise, and while what looks like a release burst is visible in the balaclava guise it is relatively low in amplitude, and spans a much smaller frequency range than the corresponding /k/ burst in the no cover guise. It is not entirely surprising, then, that several subjects should report hearing word-final /p/ in this word, given what we know about acoustic cues to place of articulation in stop consonants. Halle, Hughes and Radley (1957), for instance, specify that labial stops are cued by spectral energy predominantly at low frequencies (0.5 - 1.5kHz) while velar stops are characterised by 'strong concentrations of energy in intermediate frequency regions' (1.5 – 4kHz) (Halle *et al*. 1957: 108).

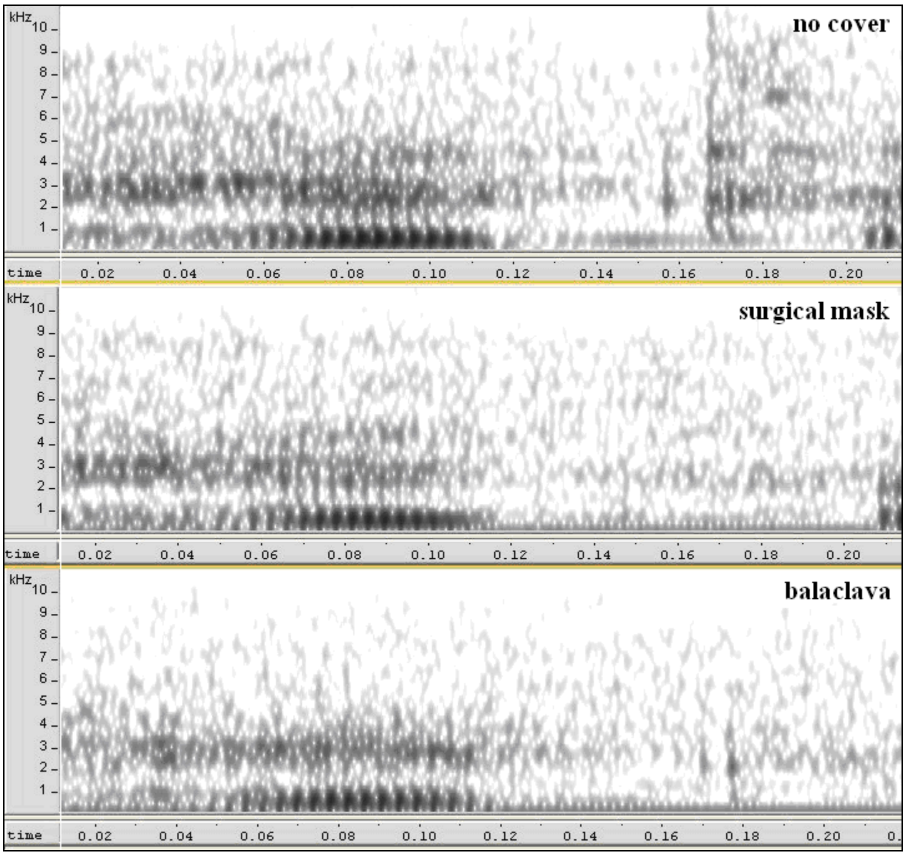

Figure 1: Wide-band spectrograms of *hick* produced by the female talker in three guises (no face covering, surgical mask, balaclava).

We have not yet systematically investigated this aspect of the perceptual strand of our experimentation with speech produced by wearers of face coverings of the sort described above, but we are confident that a significant proportion of the subjects' mishearings can be attributed relatively straightforwardly to connected speech processes such as 'accidental' voicing of intervocalic consonants, unintentional misarticulations brought about by the face covering itself (e.g., by interference with lip protrusion for certain consonants and vowels) and by the sound transmission properties of the fabric itself. It is to the last of these factors that we turn next.

4. Acoustic experiment

In order to isolate the acoustic effects of the individual face coverings, the talker and viewer-listener variables were eliminated altogether in the second experiment. Instead, the transmission loss (TL) characteristics of each of the face coverings used for the perceptual experiment, along with several other forensically-relevant fabric types for comparison, were assessed using a setup similar to that used for measuring room acoustics (e.g. Kuttruff 2000; Long 2005).

4.1. Transmission loss

Transmission loss (henceforth TL) is a property of a material that relates to its sound attenuation characteristics. TL is frequency dependent in that a greater or lesser degree of sound attenuation is observed in different parts of the frequency spectrum as a result of different energy loss mechanisms. Numerous models of TL are used in branches of industrial design and manufacture such as architectural acoustics or transport engineering, in which being able to predict the TL of materials, such as glass, brick, wood or steel, is of great importance. There exists, however, very little published research relating to TL of fabrics, including those likely to be worn as clothing, and none of these studies is specifically associated with the transmission of speech sounds.

In a study of the sound transmission characteristics of 44 fabric types, Nute and Slater (1973) evaluated the relationship between TL and fabric thickness, weight and *cover* (weave density; see Adanur 2001: 362). Fabric samples were interposed between a loudspeaker and a microphone placed at either end of a sound-wave tube, and a constant-rate sweep tone was generated for playback through the loudspeaker. Using this experimental setup, Nute and Slater found that of the three fabric-construction factors, cover had the most significant influence on TL. They also found that if it were possible to hold weight constant while increasing fabric thickness, an inverse relationship would be obtained between TL and thickness owing to the necessary increase in fabric porosity that would follow.

Porosity is also implicated in the earlier research on fabric sound absorption by Aso and Kinoshita (1963a,b; 1964), in that the size of the pores in a woven fabric relative to the wavelength of a sound signal is a factor that should be considered when assessing a fabric's TL. Aso and Kinoshita describe three mechanisms for sound absorption by fabrics: (1) viscosity-resistance, brought about by friction between air molecules and the walls of the fabric pores; (2) resonance, which is common among thicker fabrics with tightly-woven or reflective surfaces, and among fabrics under tension; and (3) a combination of the preceding mechanisms, in which viscosity-resistance is observed at higher frequencies, and resonance at lower frequencies.

A more recent study by Noy (2003) investigated three lightweight fabrics, one of which is marketed as 'acoustically transparent', and is of the sort used to cover or conceal loudspeakers. Noy found that all three fabrics reduced the amplitude of a signal when they were placed between a loudspeaker and a microphone, and concludes that 'It is immediately clear that there is no such thing as "acoustically transparent fabric"... it is safe to state that the use of any fabric covering will always decrease the quality of the sound transmission, mainly above the 10kHz frequency band – the sound will lose some of its brilliance and freshness.' Noy reports negative TL (i.e., transmission gain) at certain frequencies, a phenomenon also clearly in evidence for some fabrics in Nute and Slater's data. This might at first glance seem counterintuitive, but if one considers the fabric as an interactive element in the loudspeaker-air-microphone system rather than a simple blocking or attenuating element between the loudspeaker and microphone, it becomes apparent that the resonant characteristics of the fabric could lead to an increase in amplitude at certain frequencies.

4.2. Methods

The transmission loss of a material cannot be measured directly. Instead, the frequency response of a system, such as a microphone-loudspeaker combination, must be measured both with and without the material interposed, and then the difference between the responses is calculated. It is this difference which corresponds to the transmission loss of the material. In this experiment, a loudspeaker and a measurement microphone were placed in close proximity (50cm apart) in an acoustically treated recording booth (see Figure 2). The Maximum Length Sequence (MLS; see e.g. Rife and Vanderkooy 1989) method was employed to determine the impulse response of the microphone-loudspeaker system both with and without the fabric samples interposed. The frequency response of the system is easily determined by computing the Fast Fourier Transform (FFT) of the impulse response. The MLS method involves replaying a pseudo-random sequence of pulses generated by a computer via the loudspeaker which are then picked up by the microphone connected to the same computer. The microphone signal is processed within the computer and the impulse response is derived. This was done using PureBits *Sample Champion* v.3.8.4.0 running on a Windows PC in the recording booth control room using the default settings recommended in the manual. The loudspeaker, a Wharfedale hifi loudspeaker with three vertically-configured driver units (woofer, tweeter, woofer), was placed on the edge of a wooden table covered with a thick layer of felt to reduce vibration and acoustic reflections. The loudspeaker's position was marked on the tabletop and its base was secured to the tabletop with adhesive tape in case of accidental movement while the different fabric samples were being attached to it. The microphone, a Behringer ECM8000 omnidirectional measurement microphone, was mounted on a microphone stand, the position of which was kept constant throughout.

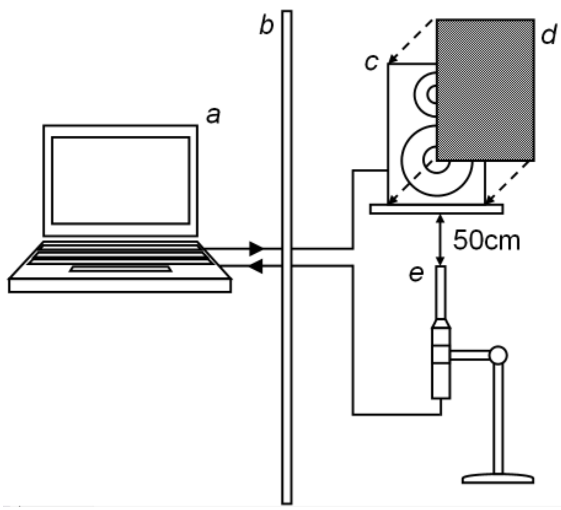

Figure 2: Experimental setup for transmission loss experiment. *a* = control room PC running Maximum Length Sequence generation and processing software; *b* = soundproofed partition wall; c = loudspeaker; *d* = fabric sample attached to clip-on frame; *e* = measurement microphone.

Some of the fabric samples were affixed to the detachable frame to which the 'acoustically transparent' fabric that had originally screened the loudspeaker driver had been stapled during its manufacture. The frame was designed to be held in place with plastic lugs, which allowed it to be removed easily and to be reattached in exactly the same position between each experimental run. The fabrics were held in place on the frame using rubber bands placed at the top and bottom (i.e. well above and below the driver units) or, in the case of the surgical mask, with adhesive tape. Note that because of the small size of the surgical mask it was necessary to attach to the frame several identical masks one above the other with as little overlap as possible. For each mask the sleeve containing a wire insert that allows the upper edge of the mask to be shaped to the bridge of the wearer's nose was removed with scissors. No attempt was made to open out the pleats on each mask, as would occur when the mask is placed over the wearer's mouth. Where it was not possible to easily attach the fabric to the frame, it was either stretched over the entire loudspeaker or draped over it so the drivers were covered.

The fabrics assessed and their method of attachment to the loudspeaker are listed in Table 4.

Fabric	**Composition**	**Attachment**
niqāb	woven polyester	frame
balaclava	knitted acrylic	stretched
disposable surgical mask	pleated paper	frame
handkerchief	woven cotton	frame
scarf	knitted wool/acrylic mix	draped
fleece material	knitted polyester ('Polartec')	draped
nylon stockings	10-denier sheer nylon	stretched
loudspeaker cover fabric	woven 'acoustically transparent' man-made fibre	frame

Table 4: Fabric types used in transmission loss experiment.

The first run of the experiment was performed without a fabric sample intervening between the microphone and the loudspeaker, which served as the control condition. The fabrics were then tested one by one, with control condition runs carried out between each fabric run. In total, then, there were 9 control condition runs. It was found that the variability that existed between the frequency responses for each of the 9 runs was negligibly small (the individual standard deviations varied by at most 1.2% of the averaged standard deviation) and we are therefore henceforth using the averaged curve of the 9 runs to represent the control condition.

We introduced one further extreme condition to provide a greater range of variability in the test conditions. This involved one of the experimenters (DW) standing between the loudspeaker and the microphone. This condition is labelled *human body* in the graphs shown in the following section.

4.3. Results

The results contradicted our expectations in that they showed there to be almost no difference between the control condition and seven of the fabrics tested. That is, interposing samples of these fabrics between the loudspeaker sound source and the measurement microphone resulted in only very small changes in the frequency responses for these fabrics. The exception to this was the surgical mask, which deviated markedly from the control condition between approximately 2.5kHz and 12.5kHz, and again between 14kHz and the high-frequency cutoff at 24kHz. A subset of the results is shown in Figure 3. The data for all fabrics except the *niqāb*, the woollen scarf, and the surgical mask have been omitted for clarity, in view of the fact that the response curves for the other fabrics agree so closely with that for the control condition. The curve for the experimenter's body test condition is also included for comparative purposes.

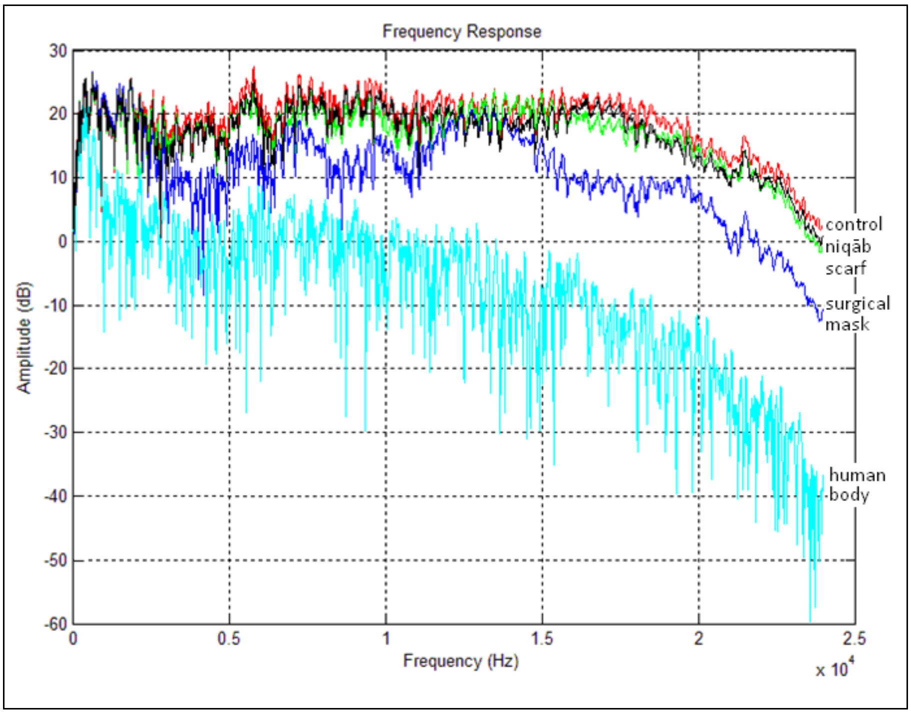

Figure 3: Frequency response curves for mean of 9 runs of control condition and 4 test conditions (*niqāb*, surgical mask, scarf, experimenter's body).

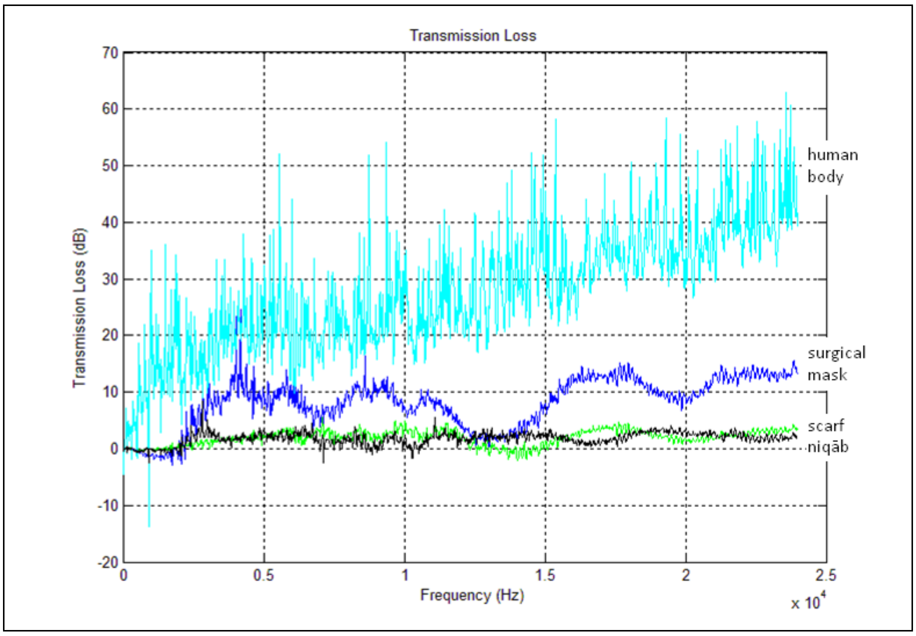

Figure 4: Transmission loss differences between frequency response curves for mean of 9 runs of control condition and 4 test conditions (*niqāb*, surgical mask, scarf, experimenter's body).

It is easier to visualise the relationship between the curves for the three fabrics if they are expressed as TL, i.e. differences from the control condition, as per Figure 4. The zero line in Figure 4 indicates parity with the control condition at any point in the 0 – 24kHz band. In the main, the curves for the *niqāb*, scarf and surgical mask, and for the experimenter's body condition, lie above the zero line, showing a loss in amplitude. It will be seen that the *niqāb* and scarf are little different from each other, and that their TL is in general quite minor. As in Figure 3, the surgical mask curve reveals considerably more TL for this fabric than is apparent for the others.

Excursions below the zero line in Figure 4 occur, albeit infrequently, for all four conditions. They indicate transmission gain, i.e. greater signal amplitude at that frequency than was the case in the control condition, and recall the findings of Nute and Slater (1973) and Noy (2003) discussed in section 4.1.

4.4. Discussion

The two key findings of the TL experiment – that the majority of fabrics are shown to have negligible TL, and that only the surgical mask can be said to affect the acoustic signal in any significant way – have interesting implications for such legal cases as might arise in which speech produced with the mouth/nose or whole face covered is of central importance. The fact that the surgical mask appears to inhibit sound transmission between loudspeaker and microphone to a much greater extent than was the case for thicker, heavier fabrics such as the woollen scarf, the balaclava and the fleece material makes it conceivable that speech produced by an individual wearing a surgical mask might, all else being equal, be less readily audible, and perhaps thereby less intelligible, than speech produced by the same person not wearing such a face covering. It seems self-evident that clear, intelligible speech between a mask-wearer and a patient, or between medical professionals in an operating theatre, might be literally of vital importance in some circumstances, and the potential for miscommunication when both visual and acoustic cues are compromised by a mask of this sort seems a very real one.

On the other hand, it must be remembered that there are major differences between the way the TL of the surgical mask has been measured and how masks of this type would be worn by medical professionals. The fact that several flattened overlapping masks held together with adhesive tape were used to cover the gap in the clip-on loudspeaker frame is a far from direct representation of the same surgical mask as worn across the mouth and nose of an anaesthetist or surgeon, in which the pleats would be opened out to achieve proper coverage over the lips and chin. The arrangement of the masks on the frame, including the use of adhesive tape, will have altered their acoustic properties compared with a single opened out mask. However, the extent of this change cannot be easily estimated and a modified experimental design would be necessary in order to accurately determine it.

Furthermore, we should not lose sight of the fact that although the distances between the loudspeaker drivers, the fabric sample, and the microphone in the TL experiment were held as constant as possible, there is still an air gap between the drivers and the fabric larger than the gap (if any) between a surgical mask-wearer's lips and the mask covering them. The same is true of any of the other fabrics worn or held against the mouth. The acoustics involved are therefore inevitably different on a range of parameters; for example, the radiation factor present in speech signals in which the talker's lip aperture opens into free air will be affected when a close-fitting mask is worn. Lastly, the fact that there is no airstream involved in the TL experiment makes the correspondences between the setup we have used and human speech production less direct still. These considerations notwithstanding, we feel we have arrived at a good first approximation of the TL properties of the various fabrics, and are confident that similar results would emerge from experiments carried out using a more realistic simulation of the talker-hearer relationship.

As for the *niqāb*, we conclude on the basis of the findings reported above that if the intelligibility of speech produced by a person wearing one is relatively difficult for listeners to understand, as prominent public figures such as Judge Glossop have claimed, the source of the problem lies not in the fabric itself but in interference to articulation caused by wearing the garment, by a disruption to or absence of visual cues available to the viewer-listener, or by some combination of these two factors. Furthermore, because the *niqāb* is a garment worn principally by women of south Asian, Afghan and Middle Eastern origin, many *niqāb* wearers in the UK may not speak English as their first language, and may perhaps have marked non-native accents that make their utterances relatively difficult for viewer-listeners to parse successfully. It does not seem improbable either that many viewer-listeners in the UK may not be familiar or comfortable with holding conversations with individuals whose faces are not fully visible, so that the tacit rules governing regular face-to-face conversational interaction between individuals with their faces uncovered can only be applied partially and awkwardly. There may also be some element of prejudice involved, such that claims of unintelligibility are made on grounds other than ones relating directly to speech perception itself, such as dislike or distrust of foreigners and/or foreign cultures.

We have thus far tested only one kind of *niqāb*. Other styles in a wide range of fabrics are available, and it is possible that heavier and/or thicker fabrics would produce different results. Given that testing a thickly knitted woollen scarf and balaclava in the present experiment did not yield TL results that were noticeably different from those observed for the *niqāb*, however, it is not obvious that this would necessarily be the case. One must, of course, bear in mind the complex relationship between thickness, weight and cover discussed by Nute and Slater (1973). It does not follow automatically that thicker, heavier fabrics have greater TL than thinner ones, if the degree of porosity and the size of the pores themselves is related to the relative levels of sound attenuation brought about by viscosity-resistance and resonance, as per Aso and Kinoshita's (1963a,b; 1964) models. Much may also depend on the degree of tension under which the fabric sample is held. Varying the type and size of the loudspeaker used might also produce rather different results, as would varying the distance between the loudspeaker and the microphone, and the position of the fabric sample between them.

5. Directions for further research

In the present context, the remarks made in the preceding section about the insignificance of the TL for certain fabrics are valid only as long as the fabrics' effects are imperceptible to the average (viewer-) listener. The next step in our research on this topic will be to assess the extent to which effects on the sound signal created by the fabric samples are audible in listening tests. This will be achieved by filtering a range of studio-quality speech recordings in *Matlab* using filters with frequency responses derived from the TL characteristics obtained for the various fabric conditions described in the preceding sections, for playback to a panel of listeners. The listeners will be asked to judge whether the second of two adjacent samples presented in randomly ordered pairs (*niqāb* condition filtered vs. unfiltered; balaclava condition filtered vs. handkerchief condition filtered, etc.) sounds different from the first, and if so whether it is, for example, 'better' or 'worse' in terms of sound quality. Response latency will also be measured. Our preliminary impressions on the basis of sound files prepared using the technique described above suggest that it would be difficult or impossible for listeners consistently to distinguish samples filtered using any of the fabric frequency response curves other than that obtained in the surgical mask condition. The application of the latter to a test sound signal (a female American English speaker reading the first line of *The North Wind and* *the Sun*) results in a duller, less 'trebly' sound quality noticeably different from that heard in the other fabric condition-filtered signals. Spectrographic comparison by eye shows that vowel formants and fricative energy above around 3.2kHz are markedly lower in amplitude in the surgical mask-filtered signal than is the case in either the unfiltered control condition-signal or the *niqāb* condition-filtered signal, confirming our auditory judgments. However, there is also visual evidence of transmission gain in vowel formants lying in the 0.5 – 2.5kHz band in the *niqāb* condition-filtered signal relative to the unfiltered control condition, and we should not rule out the possibility that even very slight differences in the lower frequency bands of the spectrum might be sufficiently auditorily salient to affect listeners' judgments. More systematic testing might yield above-chance results that suggest that listeners can react to auditory differences that are, on the face of it, so small as to be unnoticeable.

As mentioned in section 3.8, we plan a number of future experiments in which parameters relating to the visual, auditory and acoustic domains will be varied. In a revised version of the perceptual experiment the test utterances read by the actors will be altered such that subjects cannot predict the identity of a word on the basis of which words have already been presented in that particular block. The number of word misidentifications, it will be remembered from section 3.7, was overall rather low, and in order to make the task more demanding of participants, as well as (arguably) more realistic from the forensic point of view, varying levels of non-speech noise will be introduced onto the video soundtrack. By attempting to counteract the observed ceiling effect in this way, any adverse effects on speech intelligibility brought about by a face covering might become more clearly apparent.

The illumination conditions during the recording of the video film will also be modified such that the talker is lit from the side as well as from above. If the lower part of the talker's face is seen more, as it were, in 'relief', it might allow movements of the articulators to be more visible than was the case in the video shown to the Aberdeen subjects. It will be recalled that the findings of some of the audiovisual speech studies summarised in section 3.8 suggest that viewer-listeners can make use of very subtle movements of the talker's face and head in order to assist in the processing of spoken language, and even small variations in the position of a fabric covering the talker's mouth might be of perceptual relevance. The value of eye-tracking techniques will also be assessed here.

Finally, it would also be informative to run hearing tests on subjects wearing headgear which completely covers the ears, such as the *niqāb*, the balaclava, woollen or fleece hats, etc., or which occludes the ear canal in some other way. Walkman headphones (in-ear and full-ear) and hands-free telephone headsets, for instance, are now worn by large numbers of people in public spaces. From a forensic perspective, one can easily enough imagine legal scenarios in which doubt is cast on the reliability of the testimony of an earwitness to a robbery, for example, if the witness was at the time the event took place wearing a thick hat or had earphones inserted into his or her ears, even if the device to which the earphones were attached were switched off. Obtaining objective estimates of the contribution such headgear might make to impairments in word and speaker recognition by listeners, along with the implications this might have for memory for novel voices and the accuracy of recognition of known ones, seems a fruitful goal for future research.

Acknowledgments

We are grateful to the following people for advice and assistance: Fiona Ashby, Frantz Clermont, David Coniam, Bronwen Evans, Peter French, Gemma Grant, Sarah Hawkins, Mark Jones, Heuihyun Kim, Huw Llewelyn-Jones, Mike McMahon, Dirk Noy, and Jane Stuart-Smith. We also thank an anonymous reviewer for helpful comments and suggestions.

References

Adanur, S. (2001). *Handbook of Weaving*. London: CRC Press (Taylor and Francis).

Aso, S. and R. Kinoshita (1963a). Absorption of sound wave by fabrics, part I: Absorption mechanisms. *Journal of the Textile Machinery Society of Japan (English edition) 9*, 1-15.

Aso, S. and R. Kinoshita (1963b). Absorption of sound wave by fabrics, part II: Acoustic impeadence [sic] density. *Journal of the Textile Machinery Society of Japan (English edition) 9*, 40-46.

Aso, S. and R. Kinoshita (1964). Absorption of sound wave by fabrics, part III: Flow resistance. *Journal of the Textile Machinery Society of Japan (English edition) 10(5)*, 236-241.

Berger, K. W., M. Garner, and J. Sudman (1971). The effect of degree of facial exposure and the vertical angle of vision on speechreading performance. *The Teacher of the Deaf 69*, 322-326.

Bishop, J., R.H. Bahr, and M. Gelfer (1989). Near-field speech intelligibility in chemical-biological warfare masks. *Military Medicine 164(8)*, 543-550.

Bond, Z.S., T.J. Moore, and B. Gable (1989). Acoustic–phonetic characteristics of speech produced in noise and while wearing an oxygen mask. *Journal of the Acoustical Society of America 85(2)*, 907-912.

Brancazio, L. and J.L. Miller (2005). Use of visual information in speech perception: Evidence for a visual rate effect both with and without a McGurk effect. *Perception and Psychophysics 67*, 759-769.

Campbell, R. (2008). The processing of audiovisual speech: Empirical and neural bases. *Philosophical Transactions of the Royal Society of London Series B 363(1493)*, 1001-1010.

Coniam, D. (2005). The impact of wearing a face mask in a high-stakes oral examination: An exploratory post-SARS study in Hong Kong. *Language Assessment Quarterly 2*, 235-261.

Cotton, J.C. (1935). Normal 'visual hearing'. *Science 82*, 592-593.

Erber, N.P. (1979). Auditory-visual perception of speech with reduced optical clarity. *Journal of Speech and Hearing Research 22(2)*, 212-223.

Glidden, C.M. and P.F. Assmann (2004). Effects of visual gender and frequency shifts on vowel category judgments. *Acoustics Research Letters Online 5(4)*, 132-138.

Green, K.P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd, and D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, pp. 3–25. Hove: Psychology Press.

Halle, M., G.W. Hughes, and J.-P.A. Radley (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America 29(1)*, 107-116.

Hay, J., P. Warren, and K. Drager. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics 34(4)*, 458-484.

Johnson, K., E.A. Strand, and M. d'Imperio (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics 27(4)*, 359-384.

Jones, J.A. and D.E. Callan (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. *Neuroreport 14(8)*, 1129-1133.

Jordan, T.R. and P.C. Sergeant (1998). Effects of facial image size on visual and audio-visual speech recognition. In R. Campbell, B. Dodd, and D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, pp. 155-176. Hove: Psychology Press.

Judicial Communications Office, News Release 29/06. *Guidance on the wearing of veils by representatives in courts and tribunals*. URL: <http://www.judiciary.gov.uk/publications_media/media_releases/2006/2906.htm>. Accessed 13.6.08.

Kuttruff, H. (2000). *Room Acoustics, 4th edn*. London: Taylor and Francis.

Lansing, C.R. and G.W. McConkie (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. *Journal of Speech and Hearing Research 42*, 526–539.

Long, M. (2005). *Architectural Acoustics*. London: Academic Press.

Massaro, D.W. (1987). *Speech Perception by Ear and Eye: A Paradigm for Psychological Inquiry*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Massaro, D.W. (2001). Speech perception. In N.M. Smelser, P.B. Baltes (Eds.), and W. Kintsch (Section ed.), *International Encyclopedia of Social and Behavioral Sciences*, pp. 14870-14875. Amsterdam: Elsevier.

McGurk, H. and J. McDonald (1976). Hearing lips and seeing voices. *Nature 264*, 746-748.

Miller, G.A. and P.E. Nicely (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America 27*, 338-352.

Munhall, K.G. and E. Vatikiotis-Bateson (1998). The moving face during speech communication. In R. Campbell, B. Dodd, and D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, pp. 123-139. Hove: Psychology Press.

Noy, D. (2003). *Evaluation of Transmission Loss Induced by Stretched Fabric Treatments*. URL: <http://www.wsdg.com/dynamic.php5?id=resources/technology/fabric>. Accessed 13.6.08.

Nute, M.E. and K. Slater (1973). The effect of fabric parameters on sound transmission loss. *Journal of the Textile Institute 64(11)*, 652-658.

Remez, R.E. (2005). The perceptual organization of speech. In D.B. Pisoni and R.E. Remez (Eds.), The Handbook of Speech Perception, pp. 28-50. Oxford: Blackwell.

Rife, D. and J. Vanderkooy (1989). Transfer function measurement with Maximum-Length Sequences. *Journal of the Audio Engineering Society 37(6)*, 419-443.

Rosenblum, L.D. and H.M. Saldaña (1998). Time-varying information for visual speech perception. In R. Campbell, B. Dodd, and D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, pp. 61-81. Hove: Psychology Press.

Schwartz, J.-L., J. Robert-Ribes and P. Escudier (1998). Ten years after Summerfield: A taxonomy of models for audio-visual fusion in speech perception. In R. Campbell, B. Dodd, and D. Burnham (Eds.), *Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech*, pp. 85-108. Hove: Psychology Press.

Strand, E.A. and K. Johnson (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, Bielefeld, pp. 14-26. Berlin: Mouton de Gruyter.

Sumby, W.H. and I. Pollack (1954). Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America 26(2)*, 212-215.

Summerfield, A.Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd and R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lipreading*, pp. 3-51. Hove: Lawrence Erlbaum Associates.

UK Judicial Studies Board (2007). *Equal Treatment Bench Book, Chapter 3: Religious Dress*. URL: <http://www.jsboard.co.uk/downloads/ettb_veil.pdf>. Accessed 13.6.08.

Appendix A. Participant answer sheet [first page only]

Subject No.: Sex: 1st Language: Age:

Watch the screen and listen carefully. You will hear the phrase 'say X again' several times. Enter the word you hear in place of X in the appropriate space.

01	1	2	3	4	5
	6	7	8	9	10
	11	12	13	14	15
	16	17	18	19	20
	21	22	23	24	25
	26	27	28	29	30
	31	32	33	34	35
	36	37	38	39	40

02	1	2	3	4	5
	6	7	8	9	10
	11	12	13	14	15
	16	17	18	19	20
	21	22	23	24	25
	26	27	28	29	30
	31	32	33	34	35
	36	37	38	39	40

03	1	2	3	4	5
	6	7	8	9	10
	11	12	13	14	15
	16	17	18	19	20
	21	22	23	24	25
	26	27	28	29	30
	31	32	33	34	35
	36	37	38	39	40

04	1	2	3	4	5
	6	7	8	9	10
	11	12	13	14	15
	16	17	18	19	20
	21	22	23	24	25
	26	27	28	29	30

Appendix B. Participant questionnaire.

Please rank the different guises from 1 to 4, 1 being the easiest to understand and 4 being the most difficult. You can assign more than one guise the same number if you feel the intelligibility of speech was the same.

No covering ☐

Surgical Mask ☐

Veil ☐

Balaclava ☐

Do you think any other factors affected how you perceived the speech produced? You can mention something affecting particular sections or the whole recording.

A list of all of the test words follows. Please circle any you are not familiar with.

pat	tat	cat	pip	tip	kip
sip	ship	zip	seat	sheet	fin
thin	fought	thought	pick	tick	kick
Vic	mat	mats	mass	mash	match
math	Madge	hip	hat	hick	hot
heat	hit	him	kin	Kim	king
nit	mitt	vat	that		

Appendix C. Results of perception test. The criteria used for inclusion of misperceived words are discussed in section 3.7.

	male							**female**						
	audio-visual			**audio only**				**audio-visual**			**audio only**			
target word	**surg. mask**	***niqāb***	**balaclava**	**no cover**	**surg. mask**	***niqāb***	**balaclava**	**surg. mask**	***niqāb***	**balaclava**	**no cover**	**surg. mask**	***niqāb***	**balaclava**
pat	path path tat			path	paths path		path							
tat										that that			that	Pat that
cat	Cath												pit	
pip								pick						
tip									pip pip pip				pip tick	tick
pick		Vick	thick				thick			pig big			kick	big big big big big big big
tick													tip	tip
Vic	vague							big	big	big	big		big	big big Viv
fin			thin	thin	thin									
thin	fin				fin		fin				fin		fin	
fought	thought							thought				thought	thought thought	
thought								fought fought fought	forth fought	vat		fought fought	fought	
mat	math	math	mats										match	
mats	mat mat		mat		mat		mat mat mat mat mat	mat				mat		
math	mat							muff maf	maff		muff	maff	muff	muff maf
mass			mats			mat	Matt						math	
match									Madge Madge Madge				Madge Madge Madge Madge	
hip									seat					
hit	mitt									hid				hid hid
hick	heat		heat					heat hip hip hip	heat hit	hip hip		hit hip hip hip		hip hip hip hip
vat				that	that			that	that	that	that	that		that that
that													vat	
kin		king	king	king		king	king	king	king	king	king	king	king	
Kim				king				king	king				kin	king
king	keen													
nit			met		mitt				mitt	mitt				mitt mitt
mitt												knit		
sip								zip		zip		zip		zip
sit	sip													
zip														sip

Carmen Llamas
Department of Language and Linguistic Science
University of York
email: cl558@york.ac.uk

Philip Harrison
Department of Language and Linguistic Science
University of York
and
JP French Associates
86 The Mount, York
email: pth102@york.ac.uk

Damien Donnelly
Faculty of Education
University of Glasgow
email: 0704950d@student.gla.ac.uk

Dominic Watt
Department of Language and Linguistic Science
University of York
email: dw539@york.ac.uk