HAL
open science

An analysis of British regional accent and contextual cue effects on speechreading performance

Amy Irwin, Michael Pilling, Sharon. M. Thomas

To cite this version:

Amy Irwin, Michael Pilling, Sharon. M. Thomas. An analysis of British regional accent and contextual cue effects on speechreading performance. Speech Communication, 2011, 53 (6), pp.807. 10.1016/j.specom.2011.01.010 . hal-00746110

HAL Id: hal-00746110

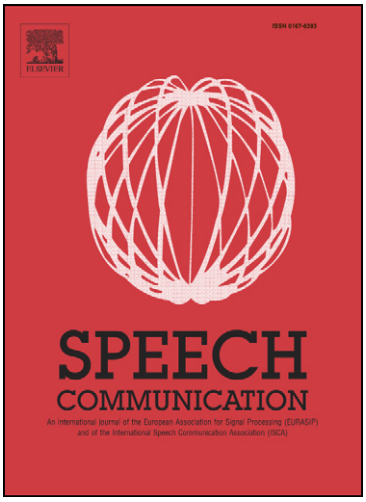
https://hal.science/hal-00746110v1

Submitted on 27 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript



An analysis of British regional accent and contextual cue effects on speech-reading performance

Amy Irwin, Michael Pilling, Sharon. M. Thomas

PII: S0167-6393(11)00017-3
DOI: 10.1016/j.specom.2011.01.010
Reference: SPECOM 1968

To appear in: *Speech Communication*

Received Date: 14 July 2010
Revised Date: 21 January 2011
Accepted Date: 26 January 2011

Please cite this article as: Irwin, A., Pilling, M., Thomas, n.M., An analysis of British regional accent and contextual cue effects on speechreading performance, *Speech Communication* (2011), doi: 10.1016/j.specom.2011.01.010

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

An analysis of British regional accent and contextual cue effects on speechreading performance.

Amy Irwin[1], Michael Pilling[1] & Sharon. M. Thomas[1]

[1]MRC Institute of Hearing Research, Nottingham

ACCEPTED MANUSCRIPT

Address for correspondence:

Dr Amy Irwin, School of Psychology, University of Aberdeen, Aberdeen, AB24 2UB

Telephone: (0044) (0)1224 273003 Fax: (0044) (0) 1224 273211

Electronic mail: a.irwin@abdn.ac.uk

ACCEPTED MANUSCRIPT

Abstract

The aim of this paper was to examine the effect of regional accent on speechreading accuracy and the utility of contextual cues in reducing accent effects.

Study 1: Participants were recruited from Nottingham (n = 24) and Glasgow (n = 17). Their task was to speechread 240 visually presented sentences spoken by 12 talkers, half with a Glaswegian accent, half a Nottingham accent. Both participant groups found the Glaswegian talkers less intelligible ($p < 0.05$). A significant interaction between participant location and accent type ($p < 0.05$) indicated that both participant groups showed an advantage for speechreading talkers with their own accent over the opposite group.

Study 2: Participants were recruited from Nottingham (n = 15). The same visual sentences were used, but each one was presented with a contextual cue. The results showed that speechreading performance was significantly improved when a contextual cue was used ($p < 0.05$). However the Nottingham observers still found the Glaswegian talkers less intelligible than the Nottingham talkers ($p < 0.05$).

The findings of this paper suggest that accent type may have an influence upon visual speech intelligibility and as such may impact upon the design, and results, of tests of speechreading ability.

Keywords: Speechreading; accent; speech perception; talker speechreadability

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1. Introduction

The influence of visual speech upon auditory speech perception is well documented. For example, the visual component of speech has been shown to facilitate auditory speech comprehension both when the auditory signal is degraded (Sumby & Pollack, 1954) and under ideal listening conditions (McGurk & MacDonald, 1976; Arnold & Hill, 2001). Visual speech is also utilised to improve speech perception when the topic of discussion is complex, or speech is produced by a talker with a strong accent (Reisberg, McLean & Goldfield, 1987). A classic demonstration of the effects of visual speech on auditory speech perception is that of the 'McGurk effect' (McGurk & MacDonald, 1976), a phenomenon in which certain combinations of incongruent visual and auditory speech tokens presented together tend to be perceived as a blending of the speech information received across the two modalities (e.g. a video sequence of a talker articulating /ga/, and an auditory recording of a talker articulating /ba/ often results in a reported percept of /da/). The McGurk effect seems to be both automatic and preattentive; it persists even if participants are made aware of the incongruence of the auditory and visual signals (Summerfield & McGrath, 1984), and has been shown to be highly resistant to experimental manipulation (e.g. alteration of face size and viewing angle; Jordan, McCotter & Thomas, 2000).

Despite the robust effect of visual speech upon auditory speech perception, the level of information that can be extracted from the visual signal alone is somewhat ambiguous. Though visual speech contains highly salient cues to certain critical aspects of speech, (e.g. place of articulation, vowel rounding), other aspects of speech (e.g. manner of articulation, voicing) have no reliable correlate that can be perceived by eye (see Summerfield, 1987). Consequently, comprehension of speech by eye alone is both more effortful than by ear and is usually far less accurate. Nevertheless, utilisation of the visual speech signal, or speechreading, is an important aid to communication with a predominantly hearing

ACCEPTED MANUSCRIPT

population, for many of the 9 million deaf and hearing-impaired individuals resident in the UK (Valentine & Skelton, 2008). Speechreading is also commonly utilised by cochlear implant and hearing-aid users to supplement their perception of audible speech (e.g. Arnold, 1997).

The importance of speechreading as a communicative tool has inspired many researchers to analyse methods of testing and improving speechreading accuracy (e.g. Auer & Bernstein, 1997). An aspect of this research has been to look at speechreading performance and the factors affecting it. There are two basic research strands which have explored this issue (Summerfield, 1992). One strand focuses on the variability in speechreading ability found across observers and has attempted to identify the factors which tend to be associated with good speechreading (Conrey & Gold, 2006). Here research seems to indicate working memory capacity and speed of lexical processing as the main correlates of proficient speechreading ability (Ronnberg, Andersson, Samuelsson, Soderfeldt, Lyxell & Risberg, 1999). The observers' viewing strategy has also been proposed as a factor in speechreading capability, for example; focus upon the talker's mouth as opposed to viewing the talker's entire face has been reported in association with reduced speechreading performance (Conrey & Gold, 2006). It is evident that observer factors can influence performance; however, these factors in isolation do not fully explain all variation in speechreading performance.

A second strand of research has focused on the variability of talkers themselves in terms of the intelligibility of the visual speech they produce (e.g. Yakel, Rosenblum & Fortier, 2000). Relatively little is known about why the speech of some talkers appears to be more visually intelligible than others, though several candidate factors for talker intelligibility have been suggested. These include lip shape, speech rate, and talker familiarity (Lesner, 1988; Massaro, Cohen & Gesi, 1993; Lander & Davies, 2008). At the most basic level, talkers have been shown to differ in the visible distinctiveness of the individual speech tokens

ACCEPTED MANUSCRIPT

they produce (Kriscos & Lesner, 1982). Visible speech tokens are typically referred to as *visemes* in the speechreading literature; a viseme can be thought of as a basic unit of the visual speech signal analogous to the phonemes of auditory speech. Each individual viseme usually encompasses several phonemes because of the considerable visual similarity of the mouth movements produced when articulating discrete phonetic tokens (for instance, for most talkers the articulations of the phonemes /b/, /p/, and /m/ would be visibly almost indistinguishable from one another and thus can be considered as a single viseme unit /b-p-m/ (Owens & Blasek, 1985). Research has shown that different talkers often vary significantly in the number of distinct visemes that can be identified from their speech. Such differences are found even within groups of talkers who have the same dialect and the same patterns of articulation (Kriscos & Lesner, 1982; 1985). This factor has a direct impact upon talker intelligibility: talkers who have a high number of distinct viseme categories in their visual speech tend to produce speech which is more visibly intelligible compared to talkers whose viseme number is more limited (Kriscos & Lesner, 1982). Finally, a combined study of the relative importance of both talker and observer variation on speechreading performance was conducted using generalizability analysis. The results estimated that the talker tends to account for 4.9% of variance, and the speechreader 10.5% of the variance (Demorest & Bernstein, 1992). There is also some evidence of subject-by-talker interactions (Demorest & Bernstein, 1992) whereby specific talkers may be less intelligible than others to only a proportion of observers, the reminder finding the same talker highly intelligible.

Auditory research has identified three main categories of auditory talker variation; i) variation associated with group membership, such as regional accent and/or dialect, ii) variation associated with the physical characteristics of the talker, such as age, gender and vocal tract shape, iii) variation associated with affective properties, such as emotional state (Abercrombie, 1967). There is also a body of research to suggest that the alteration of a

ACCEPTED MANUSCRIPT

talker's speech pattern, such as that caused by the production of 'clear speech' can affect the intelligibility of speech (Ferguson & Kewley-Port, 2002; Gagne, Masterson, Munhall, Bilida & Querengesser, 1994). This research on auditory speech suggests several possible factors which might influence the intelligibility of a talker's speech when perceived by *eye*, but which have not yet been investigated in the visual speech perception literature. In this paper we focus on one of those factors; talker accent and an observers' familiarity with that accent. Accent is a group factor, in that the characteristics of an accent type are shared across a defined group of individuals, i.e. a group of individuals born in the same geographic location will share characteristics of the same accent type. To our knowledge, the current paper represents the first examination of group effects on speechreading performance, analysed using accent as the determining group factor.

Accent has already been shown to be an important variable in auditory speech perception: When individuals listen to auditory speech produced in an accent that is unfamiliar to them, their perception of that speech is usually poorer than when listening to speech produced in an accent that they are familiar with. Indeed, lack of familiarity with an accent type has been shown to affect auditory speech comprehension in children (Nathan, Wells, & Donlan, 1998; Nathan & Wells, 2001), adults (Labov, 1989; Labov & Ash, 1997; Munro & Derwing, 1995) and upon speech in noise intelligibility (Clopper & Bradlow, 2006). Additionally, a study examining lexical decision–making (Floccia, Girard, Goslin & Konopczynski, 2006) found that auditory speech spoken in an accent which is unfamiliar to the participants (based on their geographical origin) was associated with significantly increased reaction times. Importantly it was this *familiarity with the accent* which most influenced performance; none of the accents used were shown to be inherently easier to perceive.

ACCEPTED MANUSCRIPT

As accent is an important factor in auditory speech perception it seems reasonable to ask whether accent is also important in the perception of visual speech. This is particularly true because of the many differences in the production of English across Britain, particularly across regions, each of which serves to change the manner of production and sound of speech (Wells, 1982a). For example, regional accent variation encompasses *realisational* differences; i.e. differences in the phonetic realisation (production) of speech sounds. Thus, two accents may differ in the relative length of a spoken vowel sound, or whether that vowel is produced as a monophthong (single pure sound, one tongue position) or a diphthong (two sounds blended together, two tongue positions). These realisational differences come in many forms and are responsible for much of the rhythmic diversity present within the accents of Britain (Wells, 1982b). *Lexical-incidental* accent differences, on the other hand, relate to the choice of phonemes within lexical items. That is, many of the pronunciation differences which allow observers to identify accent types often relate to the contrasting use of phonemes between accent types. For example, the pronunciation of the word 'bath' will incorporate the vowel sound /æ/ in some accents and /ɑː/ in others. Since regional accent differences alter the phonetic realisation of speech sounds, accent variation is likely to also affect factors such as viseme distinctiveness, and potentially talker speechreadability. However, to date, there is only one empirical study in the visual speech literature that has mentioned accent effects (Ellis, MacSweeney, Dodd & Campbell, 2001). This study reports a difference in performance between North and South English participants when speechreading a talker with a Southern accent. However, accent was not the principle focus of the study and as a result the evidence is anecdotal and the evaluation of an accent effect post-hoc.

In this paper we directly test the hypothesis that accent is an important variable in speechreading through two linked studies. The first study attempts to distinguish the locality of potential accent effects by investigating the impact of accent familiarity upon visual

ACCEPTED MANUSCRIPT

speech perception. In the second study we attempt to determine if it is possible to reduce accent effects through the utilisation of contextual constraints. Since accent is known to affect the prosodic, as well as phonetic, aspects of speech production (Floccia et al, 2006; Wells, 1982a) both studies involve speechreading of sentences rather than words in order to maximise the chances of finding robust effects of accent.

1.1 Study 1: The effect of regional accent type and accent familiarity on talker speechreadability

Participants from Glasgow and Nottingham, were presented with silent video recordings of talkers who themselves were either from Glasgow or Nottingham and identifiably spoke with the accent of that region. We compare the performance of these two groups of participants in speechreading talkers with these different accents. If accent is an important factor in talker speechreadability we would expect to see one of two results. First, if the findings within auditory speech research are replicated and familiarity with an accent type is the determining factor, then observers should find the talkers with their own 'home' accent easier to speechread than those with the less familiar accent. Second, if certain characteristics of an accent type directly impact speech production, it is possible that a shared accent may define some of the variation in visual intelligibility across a group of talkers with that accent. Therefore, if a particular accent type has any kind of detrimental effect upon visual intelligibility we would expect to see a difference in group intelligibility between the two accent types irrespective of familiarity.

1.1.1 Method

Participants.

Forty-one participants (age range: 18-50) were recruited for the study, all were native English speakers and reported good hearing and normal (or corrected to normal) vision. Of

ACCEPTED MANUSCRIPT

these 24 were recruited from Nottingham and 17 from Glasgow. As part of the inclusion criteria for the study participants had to be highly familiar with the accent of their region. To guarantee this, participants were screened to ensure that they had been born in their relevant region and had lived there for a minimum of four years.

Stimuli.

Stimuli consisted of 260 video clip recordings of a talker reciting sentences from the Bamford-Kowal-Bench sentence battery (Bench, Kowal & Bamford, 1979). The sentences are relatively short and simple in structure, e.g. 'Mother baked a cake' and have been utilised in several studies of speechreading performance (e.g. MacLeod & Summerfield, 1987). Video clips of 20 different BKB sentences recorded using a talker with a Southern English accent was used as a basic speechreading measure. The remaining 240 sentences were spoken by one of twelve talkers, six (3 male, 3 female) with a Nottingham accent and six (4 male and 2 females) with a Glaswegian accent. The utilisation of multiple talkers was an attempt to compensate for expected variation between individual talkers, as indicated by past research (Kriscos & Lesner, 1982; 1984). Talker variation encompasses several aspects of a talker's physiology and mode of speech production (Lesner, 1988; Massaro, Cohen & Gesi, 1993; Lander & Davies, 2008). Although we expect accent to impact talker articulation, the level of impact will vary across talkers due to the other aforementioned aspects of variation. By using multiple talkers we were able to compare variation within and across the groups, thus enabling us to determine if the effect of accent was reasonably consistent across the defined talker groups and not just a property of a single individual talker.

The representativeness of each accent type was assured by screening whether the auditory speech of each talker fulfilled certain predetermined criteria. For example, those talkers with the Glaswegian accent produced speech which was both rhotic and exhibited

ACCEPTED MANUSCRIPT

vowel shortening, as detailed by Aitkens Law (Wells, 1982a), that is, a vowel is phonetically short unless followed by a voiced fricative or /r/. The Glaswegian talkers also exhibited a lack of the phoneme / ʊ /, leading to the homophones pool – pull, full – fool etc. In comparison, the Nottingham accent is non rhotic and does include the phoneme / ʊ /. Furthermore, each of the Nottingham talkers exhibited phonemic vowel lengthening which is characteristic of this accent. These two accents were chosen due to the number of differences between them, and because of the prevalence of the Glaswegian accent in auditory studies of speech perception (Nathan, Wells & Donlan, 1998).

The presentation of each sentence in a Nottingham or Glaswegian accent was counterbalanced across participants. Counterbalancing was done by generating two sentence lists – half of the participants viewed List A spoken by the Glaswegian talkers, and List B, spoken by the Nottingham talkers. The remaining participants viewed the opposite of this (List A Nottingham talkers, List B Glaswegian talkers). Each list featured 120 sentences, with every talker producing 20 sentences each. In addition to this counterbalancing the order of presentation of clips was fully randomised across talkers and accents for each participant. Video clips were produced by recording the talker's full face using a video camera (Sony Digital Camcorder, DSR-200AP) placed 1.5 meters directly in front of the talker. During recordings the face was illuminated by three high power lamps placed at right angles to minimise shadowing and was presented in front of a neutral white background. In each video clip the recorded visual speech was preceded by 1s of the talker's static face in a closed mouth position, and followed by a further 1s of static footage also in the closed mouth position. Video clips were silent and contained only visual information of the spoken sentences. Video clips were presented on a plasma screen controlled by a G4 Apple Macintosh computer. Purpose written software programmes running on the Macintosh

ACCEPTED MANUSCRIPT

computer controlled the presentation and randomisation of the video clips and collected and recorded participant responses via the computer keyboard and hard drive.

Procedure.

The experiment took place in a quiet room in which each participant was seated at a table directly in front of the view screen. They were instructed that the talker that they would see on the computer screen would produce one sentence per video clip. They were told to watch each clip carefully and to attempt to identify what the talker had said and type their response into the computer. Participants were informed that if they did not understand the whole sentence, they should type in any word or words they thought that they understood. The experiment consisted of 2 sections; in Section 1, each participant completed the speechreading measure (talker with southern British accent), which consisted of 20 video clips presented in random order. In Section 2, each participant viewed the second set of sentences produced by the twelve accented talkers (Glaswegian or Nottingham accented talkers). Each participant was instructed that there was no time limit for completing the experiment. However, the average time taken to complete the study was around 45 minutes.

1.1.2 Results

Scoring was done using a loose keyword scoring system with errors in morphology ignored. Every sentence had 3 keywords, with a point awarded for each correctly identified keyword. Closely related examples such as plurals were also accepted. A participant's score therefore represents the number of correctly identified keywords within a sentence set. For the basic speechreading measure there were a total of 60 keywords across all the sentences. A similar level of performance was recorded for both participant groups on the speechreading measure (Nottingham participants: 13% keywords correct, Glaswegian participants: 11% keywords correct) this was confirmed through t-test analysis; $t(39) = 0.653$, $p > 0.05$, which

ACCEPTED MANUSCRIPT

indicated no significant difference in baseline speechreading ability for each participant group.

The sentences in the main body of the experiment contained 720 keywords in total. For each participant the percentage of correct responses was calculated across talkers (60 keywords per talker) and accent groups (360 keywords per group) in order to allow analysis of speechreading performance at both the talker and group level. The results were analysed using a mixed factorial ANOVA; accent (Nottingham, Glasgow) and talker constitute the two within-participant factors, with participant location (Nottingham or Glasgow) as the between-participant factor.

A significant main effect was found for accent, $F(1, 39)$ 37.765, $p < 0.01$, indicating that the intelligibility of the Glaswegian talkers (mean keywords correct: 9.2 = 3%) was significantly poorer than that of the Nottingham talkers (mean keywords correct: 27.2 = 7%), certainly the collected data indicates a reasonably consistent trend of low speechreading scores associated with the Glaswegian talkers, when compared to the Nottingham talkers (Figure 1). The individual talkers also had a significant effect on speechreadability; $F(5, 195)$ 17.169, $p < 0.01$, indicating, as expected, that there was variation in individual talker intelligibility within the accent groupings (Figure 1). Observation of Figure 1 indicates that the Glaswegian participants tended to find talkers G5 and G6 easier to speechread than Nottingham talkers N1, N2 and N4. The Nottingham participants exhibit a clearer pattern of performance across the two sets of talkers: Nottingham participants were consistently better across all Nottingham talkers in speechreading accuracy compared to the Glaswegian talkers.

** Insert Figure 1 about here **

Participant location had no overall effect on speechreading performance; $F(1, 39)$ 0.005, $p = 0.94$. However, there was a significant accent × participant location interaction;

ACCEPTED MANUSCRIPT

$F(1, 39) = 4.175$, $p < 0.05$ (Figure 2, Panel A). There were no further significant interactions between the factors.

**Insert Figure 2 about here **

The significant interaction depicted in Figure 2 illustrates that the main effect of accent is modified by participant location. Most noticeably, each group of participants appears to be somewhat more accurate at speechreading talkers with their own home accent than participants from the opposite group, i.e. Glaswegian participants generally exhibited greater accuracy when speechreading the Glaswegian talkers than the Nottingham participants and vice versa. Observer variation was present within both participant groups (Figure 2, Panel B), illustrating the dual impact of talker and observer based factors upon speechreading performance.

1.1.3 Discussion

The results presented for Study 1 were multifaceted and as such will be discussed point by point. First, there was a significant main effect of accent type – indicating that talker variation at the group level, defined here as accent, had a significant impact upon performance. This basic pattern was found for all participants. Specifically, it would appear that the Glaswegian speakers were generally less intelligible than their Nottingham counterparts, at least for the sample of talkers used in this study. This is the first attempt at clarifying talker speechreadability at the group level, and as such suggests that the topic of accent variation is worthy of further research.

At first glance that result is at odds with similar research utilising auditory speech (Labov, 1989; Labov & Ash, 1997; Munro & Derwing, 1995) where observer accent familiarity is the determining factor in performance, as opposed to the characteristics of the

ACCEPTED MANUSCRIPT

accent type. This was not entirely unexpected since previous research has shown that talker effects are not always entirely analogous across the auditory and visual modalities (Sheffert & Fowler, 1995) and so it was possible that accent variation might affect visual intelligibility differently. However, the significant interaction between accent and participant location suggests that the effect of accent alters across participant location. Essentially, although both sets of participants found the Nottingham talkers, as a group, more intelligible, each set exhibited an advantage for their own 'home accent' group when their results were compared, as illustrated in Figure 2 (panel A). This suggests that although the Glaswegian participants did not exhibit the expected behaviour, whereby they should have been more accurate at speechreading the Glaswegian talkers than the Nottingham talkers, they did show an advantage over the Nottingham participants when speechreading the Glaswegian talkers. It is possible that this advantage relates to the Glaswegian observers experience with the Glaswegian accent, leading them to anticipate certain visual patterns that are unfamiliar to the Nottingham participants and vice versa.

The third aspect of the presented data is the effect of individual talker variation upon the results. A degree of variation was expected given that previous research has shown talker variability effects across auditory, audiovisual and visual speech stimuli (Gagne, Masterson, Munhall, Bilida & Querengesser, 1994; Kriscos & Lesner, 1982; Cox, Alexander & Gilmore, 1987), with signal intelligibility varying significantly on a talker by talker basis in all three mediums. Thus, we anticipated that intelligibility would vary at both the individual and group level, as found in the results. It is interesting to note that the advantage exhibited by each group for their own home accent (in comparison to the other group), indicated by the significant interaction reported between accent and participant location, appears to be consistent across only 11 of the 12 talkers. The discrepancy in this pattern was for talker N3, where the Glaswegian participants tended to be more accurate at speechreading that specific

ACCEPTED MANUSCRIPT

talker than the Nottingham participants. It is difficult to explain this particular aspect of our findings, it may simply be a statistical anomaly, or may reflect some aspects of that particular talker's speech which may have made it easier for the Glaswegian group to comprehend.

Finally, it should be noted that the speechreading performance scores obtained from our participants were rather low in comparison to most other studies using equivalent (BKB) stimulus materials. For instance, MacLeod & Summerfield (1987) found average performance ranged from around 20 to 50% of keywords, compared to the less than 10% obtained in the current study. This is a relatively common problem in tests of speechreading performance, with authors utilising various methods to improve performance; such as the inclusion of a low level of auditory speech (Landers & Davis, 2008). However, specifically for this paper the low scores may be due partly to the use of talkers with regional accents (rather than talkers with Received Pronunciation; as are commonly used when recording stimulus materials for speechreading research) but also due to the presentation of multiple different talkers within a single experimental session. Speechreading accuracy tends to be reduced when performance is tested with stimuli recorded by several different talkers, compared to the presentation of a single talker across all trials (see Yakel, Rosenblum & Fortier, 2000, for a discussion of multiple talker effects).

This methodological issue of poor performance is addressed in Study 2. In this experiment contextual cues were used in an attempt to significantly improve speechreading performance to a more measureable range in order to provide a further test of the effect of accent variation on talker speechreadability.

1.2 Study 2: The effect of contextual cues and repetition on the perceived intelligibility of talkers with the Glaswegian accent

ACCEPTED MANUSCRIPT

A large body of research exists on the advantages of contextual cues in auditory speech perception. For example, context has been utilised as an aid for hearing-impaired participants, with sentences (Most & Adi-Bensaid, 2001; Grant & Seitz, 2000) or preceding questions (Flynn & Dowell, 1999) improving word comprehension when the auditory signal is degraded. Contextual cues have also been used, though with minimal success, in an attempt to alleviate the comprehension difficulties that can be caused by listening to speech in an unfamiliar auditory accent (Labov, 1989; Labov & Ash, 1997). It is likely that the reported facilitation comes about through linguistic redundancy, which relates to the use of language constraints to compensate for any gaps in perception caused by difficulty in perceiving the acoustic speech signal (Boothroyd, 1988). Essentially, context helps an observer form expectations about the content of a message and then use those expectations to constrain the possible interpretations of the speech signal and aid comprehension. However, the use of context does not appear to improve perception equally and in all cases – as illustrated by the minimal success of contextual cues in improving the perception of unfamiliar auditory accents.

The effect of context on visual speech comprehension has also been investigated, with similar results obtained to those found in auditory speech. For example, a question-answer format with an initiating orthographic utterance (question) providing context, tends to improve speechreading of visual sentences (Erber, 1992). Moreover, research indicates that context in the form of a priming word (presented as text) can lead to improved perception of spoken visual words (Lansing & Helgeson, 1995). Finally, research has also indicated that a combination of cue words (i.e. the sentence 'the tie is black' might be cued by the words 'clothes' and 'colour'; Lidestam, Lyxell & Lundeberg, 2001) was enough to improve speechreading performance for sentences. However, the authors note that the effect of

ACCEPTED MANUSCRIPT

context appeared to vary according to task difficulty; the more difficult the task the greater the improvement provided by contextual cues (Lidestam, Lyxell & Lundeberg, 2001).

Thus, in the current study, if orthographically presented contextual cues are effective at improving speeachreading performance we should see an increase in reported accuracy across all 12 talkers and both accent groups. Since contextual cues, at least in the auditory modality, seem to have limited effects upon the relative ability to comprehend an unfamiliar accent, it can be predicted that Glaswegian talkers should remain relatively less intelligible than the Nottingham talkers when viewed by observers from Nottingham. Overall, it is expected that the improvement in speechreading performance should remove the floor effects previously observed in Study 1, allowing any reported accent effects to be analysed more effectively.

1.2.1 Method

Participants

Fifteen participants were recruited for the study (age range 18 -50), all were native English speakers and reported good hearing and normal (or corrected to normal) vision. As before, participants were screened to ensure they had each been born in the East Midlands area and had lived in Nottingham for more than four years. None had taken part in Study 1.

Stimuli

Stimuli consisted of the same 240 video clips as previously described in Study 1; with recordings from 12 talkers, six from each accent type (Glaswegian and Nottingham) reciting sentences from the BKB sentence battery (Bench, Kowal & Bamford, 1979). Video clips were silent and contained only visual information of the spoken sentences. Each clip was preceded by an orthographic cue word that provided topical constraint for the sentence that appeared afterwards, e.g. the word ‘property’, presented as text, might precede the sentence

ACCEPTED MANUSCRIPT

'the family bought a house'. Video clips were presented on a plasma screen controlled by a G4 Apple Macintosh computer. The cue word was presented in the centre of the computer screen for 2s preceding presentation of the video clip. The computer also controlled the presentation and full randomisation of the video clips and collected and recorded participant responses via the computer keyboard and hard drive.

Procedure.

Participants were each seated at a table directly in front of the computer screen, and instructed that they would first be shown a 'cue word', described as providing a clue as to the nature of the sentence. After the cue word had appeared a video clip would follow shortly after. Participants were instructed that the talker would produce one sentence per video clip, which they were asked to watch carefully. Their task was to identify what the speaker had said and type their response into a keyboard. They were not required to understand the entire sentence; any word that was typed in was recorded. All fifteen subjects viewed the clips presented in a randomised order. Each participant was instructed that there was no time limit for completing the experiment. However, as before, the average time taken to complete the study was around 45 minutes.

1.2.2 Results

Scoring was done using a loose keyword scoring system with errors in morphology ignored, as described previously in Study 1. Observation of the pattern of accuracy indicated a significant decrement in speechreading accuracy associated with the Glaswegian talkers (mean keywords correct: 42.27 = 12% correct) compared against the Nottingham talkers (mean keywords correct: 64.93 = 18% correct). It should be noted that performance in both cases was generally improved when compared against the performance shown in Study 1; with scores ranging from 12 to 18% correct across accent groups in comparison to 3 to 7%

ACCEPTED MANUSCRIPT

correct from the same recordings in Study 1. The results were analysed initially using a repeated measures ANOVA with accent and talkers as the two factors. Accent type was shown to have a significant effect upon performance; $F(1,18)$ 32.362, $p < 0.01$. There was also a significant effect of talker variation upon the results; $F(5, 90)$ 17.182, $p < 0.01$ (Figure 3), but no significant interaction between the two factors; $F(5, 90)$ 2.175 $p > 0.05$.

** Figure 3 about here **

A mixed factorial ANOVA was performed to analyse the relative benefit provided by contextual cues across talkers and accent groups. This was done by combining the accuracy data from the Nottingham participants in Study 1 with the data gathered in Study 2 in a single analysis. In this analysis there were two repeated measures factors; accent and talker, and a single between-participants factor; context. The results indicate a significant effect of context upon speechreading performance; $F(1, 37)$ 30.660, $p < 0.01$ (Figure 4).

** Figure 4 about here **

There was also the expected significant effect of accent; $F(1, 37)$ 63.505, $p < 0.01$, and talker variation; $F(5, 185)$ 26.135, $p < 0.01$ on speechreading accuracy. These main effects were expected given our initial analysis of both datasets, however, there was also a significant interaction between talker and context; $F(5, 185)$ 2.665, $p < 0.05$. This indicates that the effect of context was modified across the talkers, meaning the relative increase provided by a contextual cue was not equal across all of the talkers (see Figure 3). There were no further significant interactions.

1.2.3 Discussion

The replication of a significant main effect for accent type indicates that observers from Nottingham tend to find the Glaswegian accent less intelligible than their own 'home'

ACCEPTED MANUSCRIPT

accent. The results reported here and in Study 1, when viewed together, suggest that the talker-observer interaction proposed by Demorest and colleagues (Demorest & Bernstein, 1992) is potentially at work, at least at the group level; with accent type influencing the general visual intelligibility of a set of talkers defined by accent-group, when viewed by observers with a particular accent type. This has potential ramifications for tests of speechreading performance; an accent mismatch between talker and observer during assessment of speechreading capability may have an adverse effect on measured performance. However the findings here are preliminary and further work is needed across a greater range of accents before the relative importance of accent compared to other talker variables can be determined.

Auditory research utilising context to reduce accent effects (Labov, 1989; Labov & Ash, 1997) reported similar results to those found in Study 2, with accent effects proving somewhat resistant to contextual cues. Labov (1989) suggests that the unfamiliar phonetic form can essentially 'block' accurate perception of the message. That is, even when the context given clearly indicates the correct interpretation of the stimulus, the listener is unable to ignore the unfamiliar pronunciation, leading to misinterpretation of the stimulus. A similar mechanism might account for the effects found here in speechreading.

The observed general improvement in performance found with the addition of contextual cues demonstrates the value of context as an aid to speechreading. This is consistent with several previously reported findings (Lansing & Helgeson, 1995; Lidestam, Lyxell & Lundeberg, 2001). It has been suggested that measures of visual speech perception utilising context may be a more accurate method of testing, being closer to the natural flow of conversation than sentences or words produced in isolation (Flynn & Dowell, 1999). Certainly, conversation within an everyday setting should allow the average speechreader to utilise a variety of cues, including linguistic redundancy. These cues possibly allow the

ACCEPTED MANUSCRIPT

speechreader to 'fill in the gaps' left by the less visible aspects of articulation (Boothroyd, 1988) hence the general increase in speechreading performance.

1.3 Study Limitations

The accents used here were chosen because they differed on a number of points, designated as 'regional level' differences (Wells, 1982a). However, the results show a difference in speechreadability between Glaswegian and Nottingham accents as opposed to providing evidence for accent effects as a whole. Further research utilising accents that differ on a sub-regional level (such as a comparison of Sheffield and Nottingham accents; Wells, 1982a) could help to determine the sensitivity of the visual processing system to accents which share a greater number of characteristics.

The participants used in both studies were normal hearing as opposed to deaf. This is a common method of investigating visual speech (see Arnold, 1997 for a review). However, previous research comparing the speechreading performance of both hearing and deaf participants has shown that deaf speechreaders tend to be more proficient than their hearing counter-parts (Bernstein, Auer & Tucker, 2001). It might be interesting, therefore, to evaluate the performance of deaf speechreaders when they are presented with talkers producing speech in different accent types; in order to appraise the magnitude of accent effects upon such experienced speechreaders.

Finally, the results show the difficulty of determining accent effects in speech reading performance due to individual talker variability. The strategy pursued by this paper was the utilisation of multiple talkers in order to produce a mean score for each accent group, this same method has also been utilised in auditory studies of accent (Floccia, Girard, Goslin & Konopczynski, 2006). Several alternatives to this group design could be used, but each has both advantages and disadvantages. First, a single talker could be utilised, whereby that

ACCEPTED MANUSCRIPT

talker would 'produce' each of the different accents required. However, there is a basic validity issue with that method, in that each accent the talker produces which is not their own is unlikely to be a truly accurate representation. Second, two talkers could be utilised, one representing each accent type. However, as the present results show, talkers within an accent group are unlikely to be uniform in their intelligibility, thus the talkers selected may vary in intelligibility due to factors other than their accent type, making it difficult to separate idiosyncratic talker factors from accent effects. Finally, a synthetic face could be utilised (Massarro, 1998); this would almost certainly remove some of the ecological validity inherent in using recordings of actual speakers, but would allow the comparison of multiple accents across identical talker faces. An additional benefit of such research would be a potential improvement in the effectiveness of 'talking heads' by incorporating accent variation, thereby allowing programmes such as 'Synface' (Beskow, Karlsson, Kewley & Salvi, 2004) to emulate different speakers more closely.

1.4 Conclusion

The current paper has illustrated a potential effect of accent variation on talker speechreadability. This effect was apparent as both a group factor (in that one of the tested accents –Glaswegian- seemed generally harder to speechread regardless of ones home accent) and as a factor which interacted with an observers home accent. However, accent was clearly not the only factor which influenced visual intelligibility, as both talker and observer effects were also apparent from the data. Study 2 introduced contextual cues as a possible method of reducing accent effects; the results suggest that this is a relatively ineffective method of improving the intelligibility of Glaswegian talkers in comparison to Nottingham talkers. The contextual cues used did appear to improve general speechreading performance however, adding some validation to their use as a speechreading aid. Overall, the findings of this paper suggest that future tests of speechreading ability should attempt to incorporate multiple

ACCEPTED MANUSCRIPT

talkers with a variety of accent types in their design; both to increase ecological validity and to compensate for the influence of talker factors (both at the individual and group level) upon the results.

1.5 References

Abercrombie, D. (1967) *Elements of General Phonetics,* Chicago: Aldine

Arnold, P. (1997) The structure and optimization of speechreading, *The Journal of Deaf Studies and Deaf Education, 2,* 199 - 211

Arnold, P. & Hill, F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92,* 339-355.

Auer, E.T. Bernstein, L.E. (1997) Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced phonetic distinctiveness on lexical uniqueness, *Journal of the Acoustical Society of America, 102, 6,* 3704-3710.

Bench, J. Kowal, A. & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially hearing children. *British Journal of Audiology, 13,* 108-112.

Bernstein, L.E. Auer, E.T. & Tucker, P.E. (2001) Enhanced speechreading in deaf adults: can short-term training/practice close the gap for hearing adults. *Journal of Speech, Language and Hearing Research, 44,* 5-18.

Beskow, J. Karlsson, I. Kewley, J. & Salvi, G. (2004) SYNFACE – a talking head telephone for the hearing-impaired. In K. Miesenberger, J. Klaus, W. Zagler & D. Burger (Eds), *Computers helping people with special needs,* 1178 – 1186, Paris, Springer-Verlag.

Boothroyd, A. (1988) Linguistic factors in speechreading, in, C.L. De Filippo & D.G. Sims (eds) *The Volta Review, New Reflections on Speechreading, 90, 5,* 77-87

ACCEPTED MANUSCRIPT

Clopper, C.G. & Bradlow, A.R. (2006). Effects of dialect variation on speech intelligibility in noise. *Journal of the Acoustical Society of America, 119, 5,* 3424.

Conrey, B. & Gold, J.M. (2006) An ideal observer analysis of variability in visual-only speech, *Vision Research, 46, 19,* 3243-3258

Demorest, M.E. & Bernstein, L.E. (1992). Sources of variability in speechreading sentences: a generalizability analysis. *Journal of Speech and Hearing Research, 35,* 876 – 891.

Ellis, T. MacSweeney, M. Dodd, B. & Campbell, R. (2001). TAS: A new test of adult speechreading. Deaf people really can be better speechreaders. *Proceedings from AVSP 2001 International Conference on Auditory-Visual Speech Processing,* 13-17.

Erber, N. (1992). Effects of question-answer format on visual perception of sentences, *Journal of the Academy of Rehabilitative Audiology, 25,* 113-122

Ferguson, S.H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners, *Journal of the Acoustical Society of America, 12, 1,* 259 – 271.

Floccia, C. Girard, F. Goslin, J. & Konopczynski (2006). Does regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32, 5,* 1276-1293.

Gagne, J.P. Masterson, V. Munhall, K.G. Bilida, N. & Querengesser, C. (1994) Across talker variability in auditory, visual and audiovisual speech intelligibility for conversational and clear speech, *Journal of the Academy of Rehabilitative Audiology, 27,* 135 – 158.

ACCEPTED MANUSCRIPT

Grant, K.W. & Seitz, P.F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *Journal of the Acoustical Society of America. 107, 2,* 1000 – 1011.

Jordan, T.R. McCotter, M.V. & Thomas, S. (2000) Visual and audiovisual speech perception with colour and gray scale facial images. *Perception and Psychophysics, 62, 7,* 1394-1404.

Kricos, P.B. & Lesner, A. (1982). Differences in visual intelligibility across talkers. *The Volta Review*, *84,* 219 – 225.

Kricos, P.B. & Lesner, A. (1985). Effect of talker differences on the speechreading of hearing impaired teenagers. *The Volta Review, 87,* 5-16.

Labov, W. (1989). The Limitations of Context: Evidence from misunderstandings in Chicago. *Chicago Linguistic Society, 25,* 171-200.

Labov, W. & Ash, S. (1997). Understanding Birmingham, in C. Bernsteing, T. Nunnally & R. Sabino (eds) *Language Variety in the South Revisited,* Alabama University, 508-573.

Lander, K. & Davies, R. (2008). Does face familiarity influence speechreadability? *The Quarterly Journal of Experimental Psychology, 61, 7,* 961-967.

Lansing, C.R. & Helgeson, C.L. (1995) Priming the visual recognition of spoken words. *Journal of Speech and Hearing Research, 38,* 1377-1386

Lesner, S.A. (1988). The Talker, *In* C.L. Fillipo & D.G. Sims (eds) *New Reflections on Speechreading: The Volta Review*, *90,* 5, 89-98.

ACCEPTED MANUSCRIPT

Lidestam, B. Lyxell, B. & Lundeberg, M. (2001) Speechreading of synthetic and natural faces: Effects of contextual cueing and mode of presentation, *Scandinavian Audiology, 30, 2,* 89-94.

MacLeod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise, *British Journal of Audiology, 21, 22,* 131-141.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, Massachusetts: MIT Press.

Massaro, D.W. Cohen, M.M. & Gesi, A.T. (1993). Long-term training, transfer and retention in learning to lipread. *Perception and Psychophysics, 5,* 549-562.

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

Most, T. & Adi-Bensaid, L. (2001) The influence of contextual information on the perception of speech by post-lingually and pre-lingually profoundly hearing-impaired Hebrew speaking adolescents and adults, *Ear and Hearing, 22, 3,* 252-263

Munro, M.J. & Derwing, T.G. (1995). Processing time, accent and comprehensibility in perception of native and foreign-accented speech. *Language and Speech, 38,* 289-306.

Nathan, L. Wells, B. & Donlan, C. (1998). Childrens comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language, 25,* 343-365.

Nathan, L. & Wells, B. (2001). Can children with speech difficulties process an unfamiliar accent? *Applied Psycholinguistics, 22,* 343-361.

Reisberg, D. McLean, J. & Goldfield, A, (1987) Easy to hear but hard to understand: a lipreading advantage with intact auditory stimuli, 97 - 113 *in* Dodd, B & Campbell, R. (eds)

ACCEPTED MANUSCRIPT

(1987) *Hearing By Eye: The Psychology of Lip-Reading,* Lawrence Erlbaum Associates, London

Sheffert, S.M. & Fowler, C.A. (1995) The effects of voice and visible speaker change on memory for spoken words, *Journal of Memory and Language,* 34, 665 – 685.

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audiovisual speech perception *in,* Dodd, B & Campbell, R. (eds) (1987) *Hearing By Eye: The Psychology of Lip-Reading,* Lawrence Erlbaum Associates, London.

Valentine, G. & Skelton, T. (2008). Changing spaces: The role of the internet in shaping deaf geographies, *Social and Cultural Geography, 9, 5,* 469 – 485.

Wells, J.C. (1982a). *Accents of English 1: An Introduction,* University Press, Cambridge.

Wells, J.C. (1982b). *Accents of English 2: The British Isles,* University Press, Cambridge.

Yakel, D.A. Rosenblum, L.D. & Fortier, M.A. (2000). Effects of talker variability on speechreading. *Perception and Psychophysics, 62, 7,* 1405-1412.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

1.6 Figure captions

Figure 1: Bar chart of speechreading accuracy for each of the twelve talkers who featured in the study, across participants from Glasgow and Nottingham. The scale on the *y* axis represents performance, defined as the mean percentage of keywords correctly identified for sentences spoken by each talker with a Glaswegian (G) or Nottingham (N) accent. Shown with standard error bars.

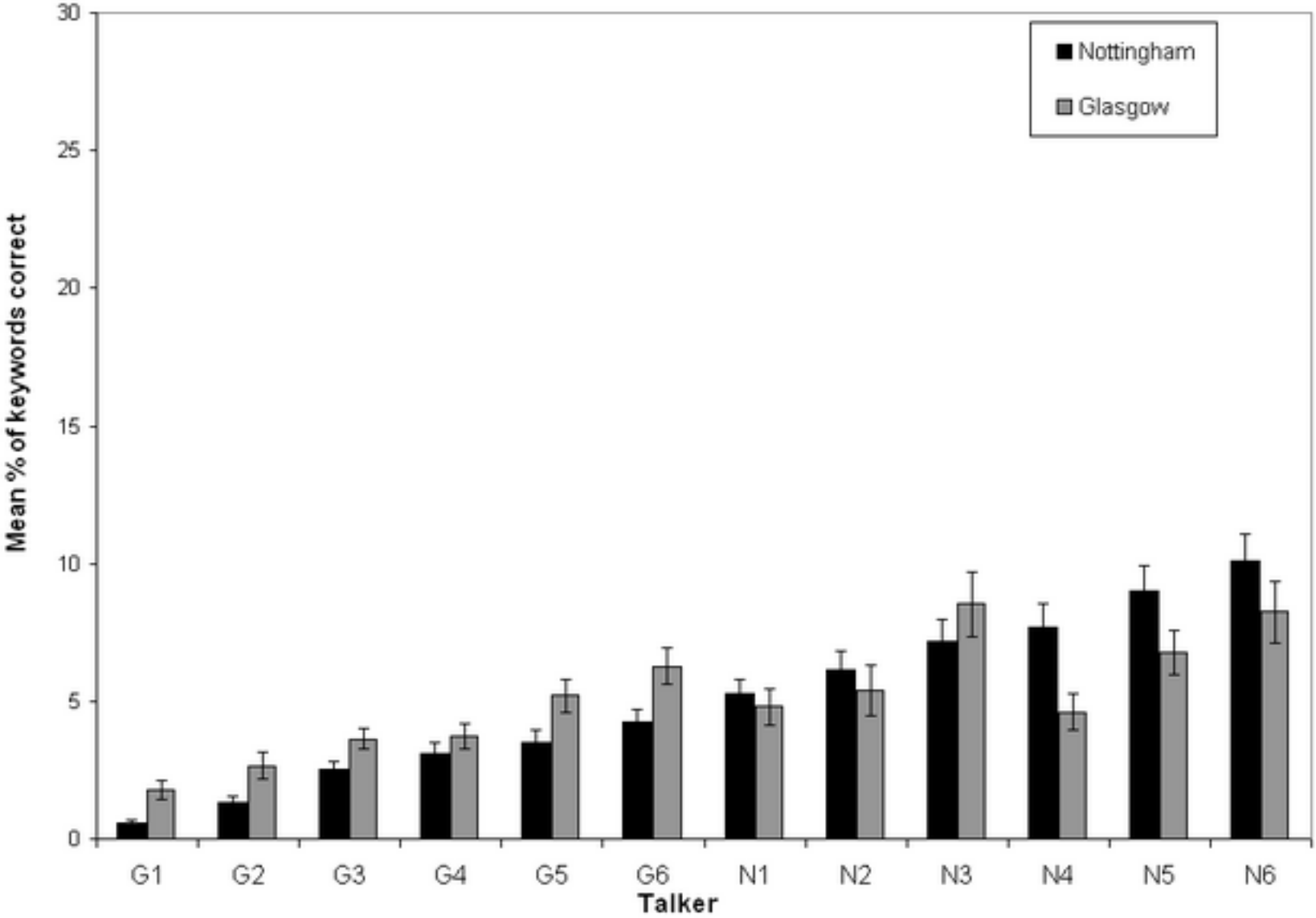
Figure 2: Panel A displays a line graph of speechreading accuracy across participants from the Glasgow and Nottingham regions, shown with standard error bars. Panel B displays boxplots of same data. The scale on the *y* axis for both panels represents performance, defined as the mean percentage of keywords correctly identified for sentences spoken by a talker with a Glaswegian or Nottingham accent.

Figure 3: Bar chart of speechreading accuracy for each of the twelve talkers who featured in Study 1 and 2, across the Nottingham participants from Studies 1 (no-context) and 2 (context). The scale on the *y* axis represents performance, defined as the mean percentage of keywords correctly identified for sentences spoken by each talker with a Glaswegian (G) or Nottingham (N) accent. Shown with standard error bars.

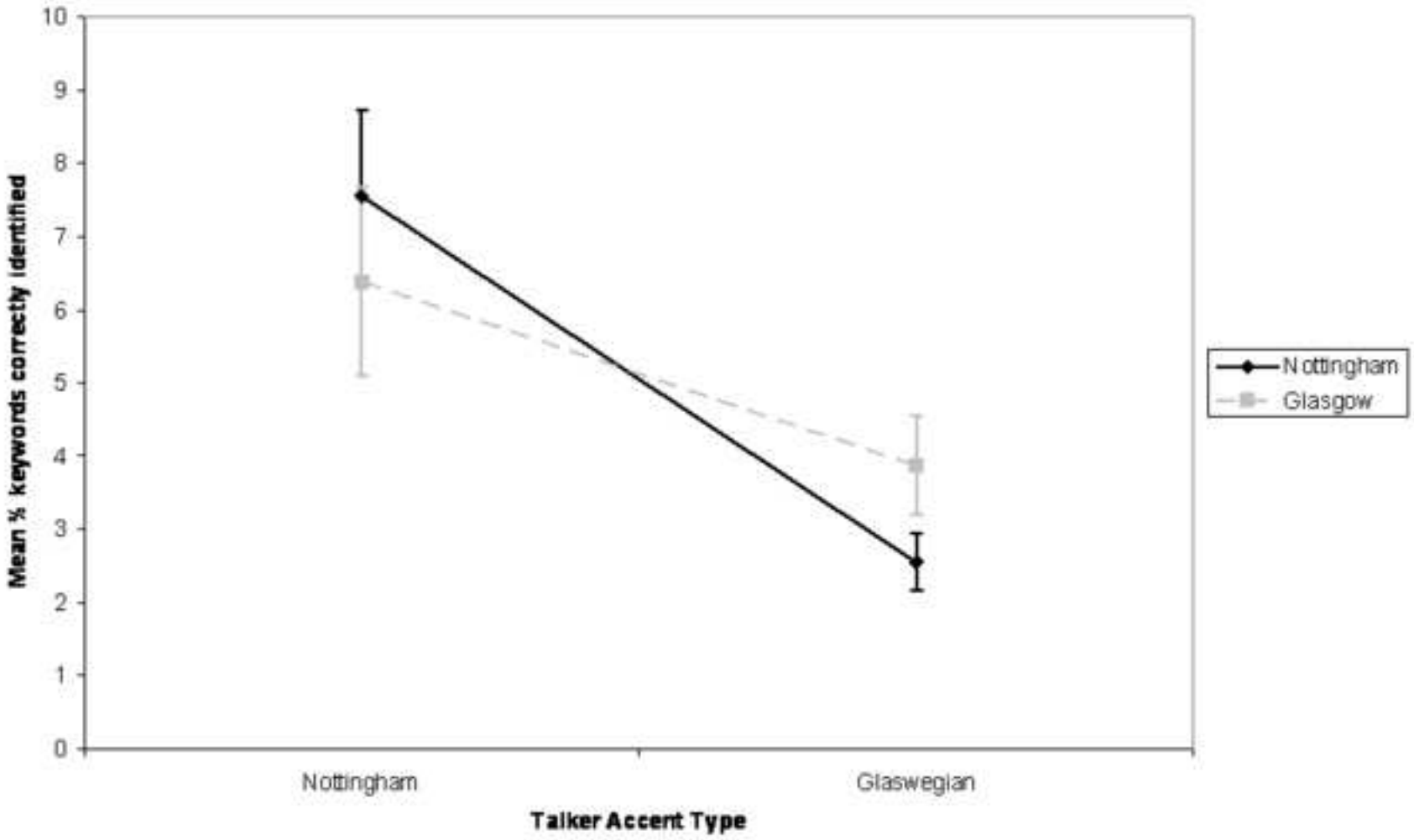
Figure 4: Bar chart of speechreading accuracy across participants presented with a contextual cue (Study 2) compared to those who viewed the video clips with no cue present (Study 1). The scale on the *y* axis represents performance, defined as the mean percentage of keywords correctly identified for sentences spoken by a talker with a Glaswegian or Nottingham accent. Shown with standard error bars.

ACCEPTED MANUSCRIPT

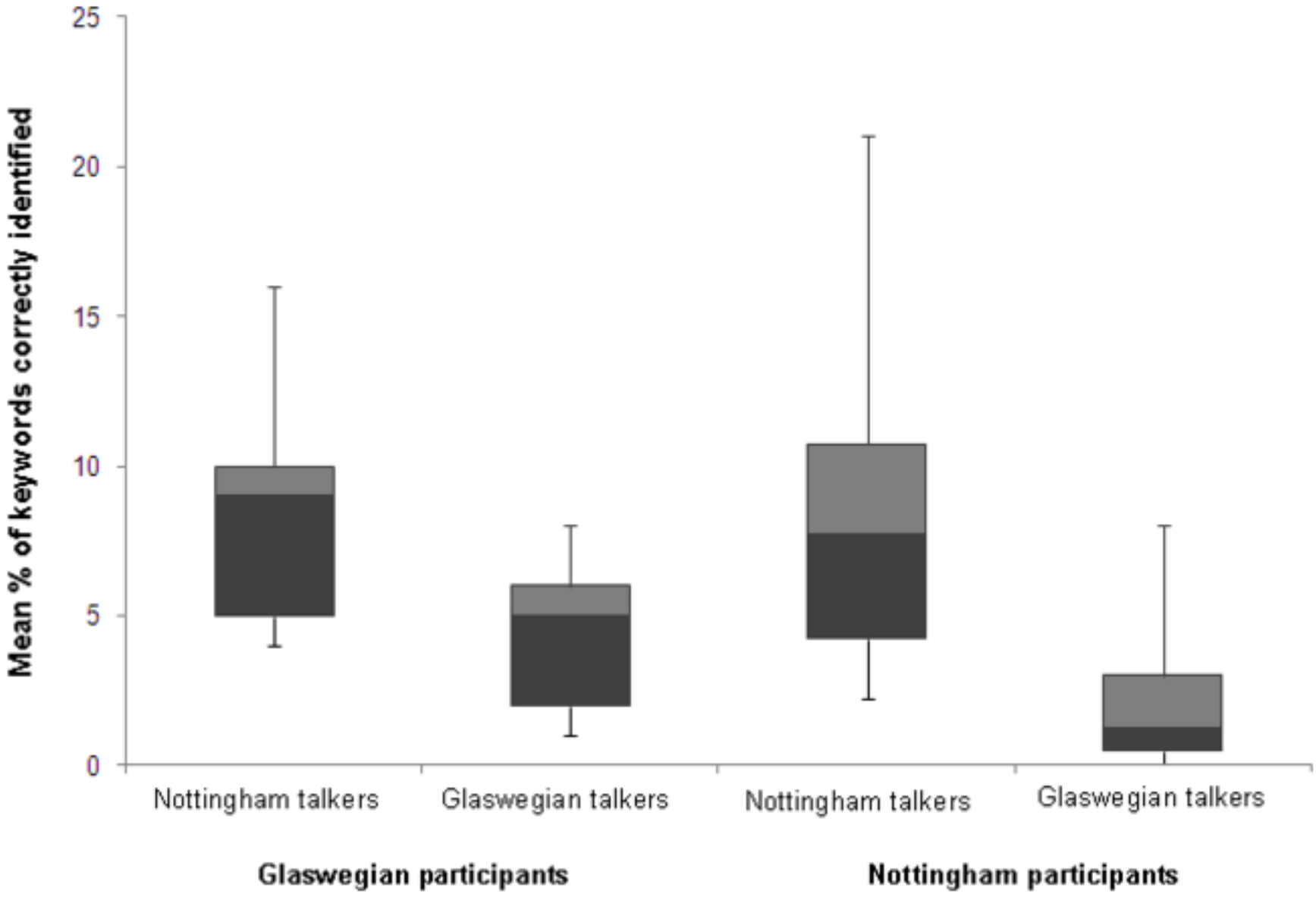
ACCEPTED MANUSCRIPT



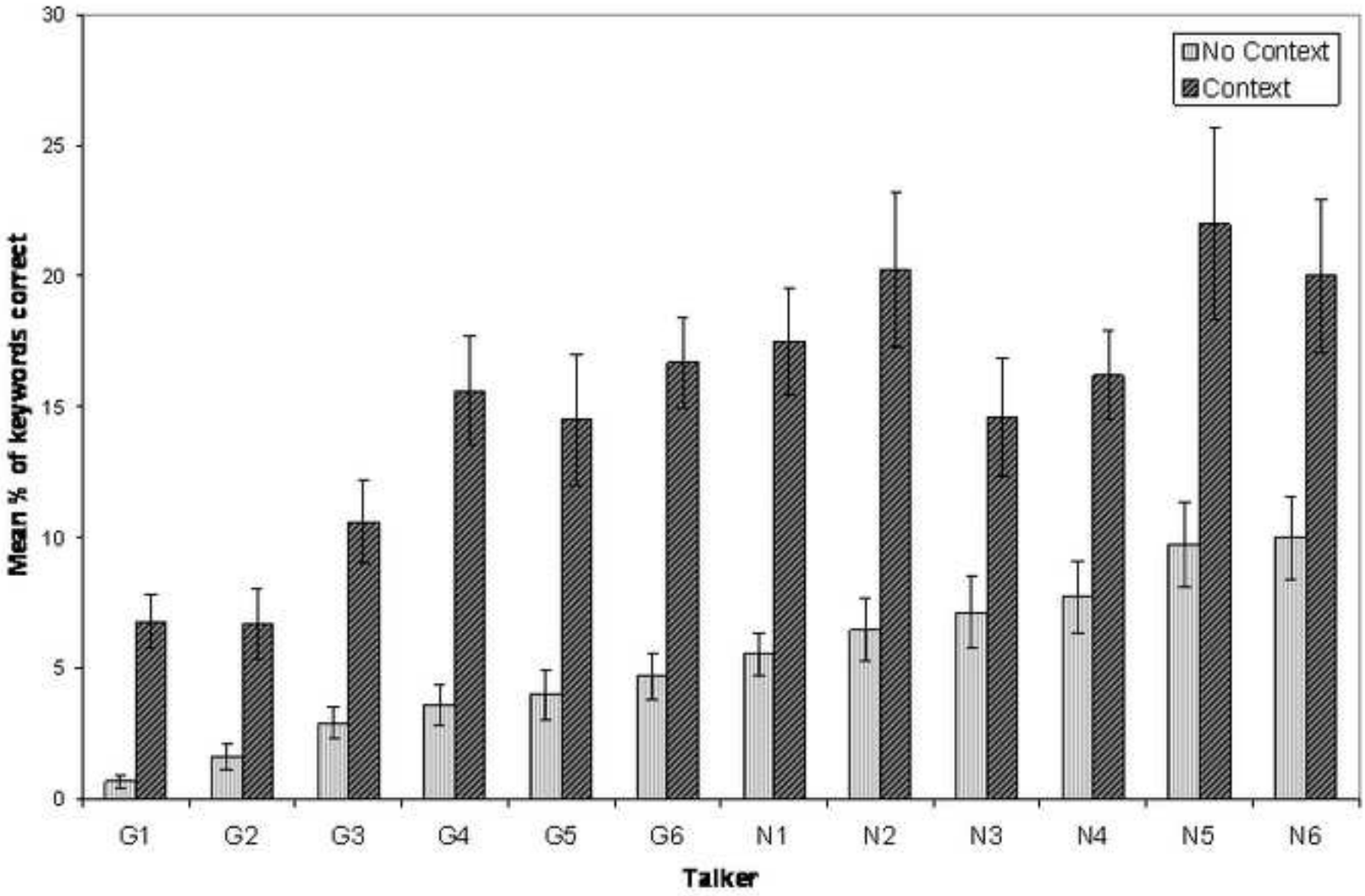
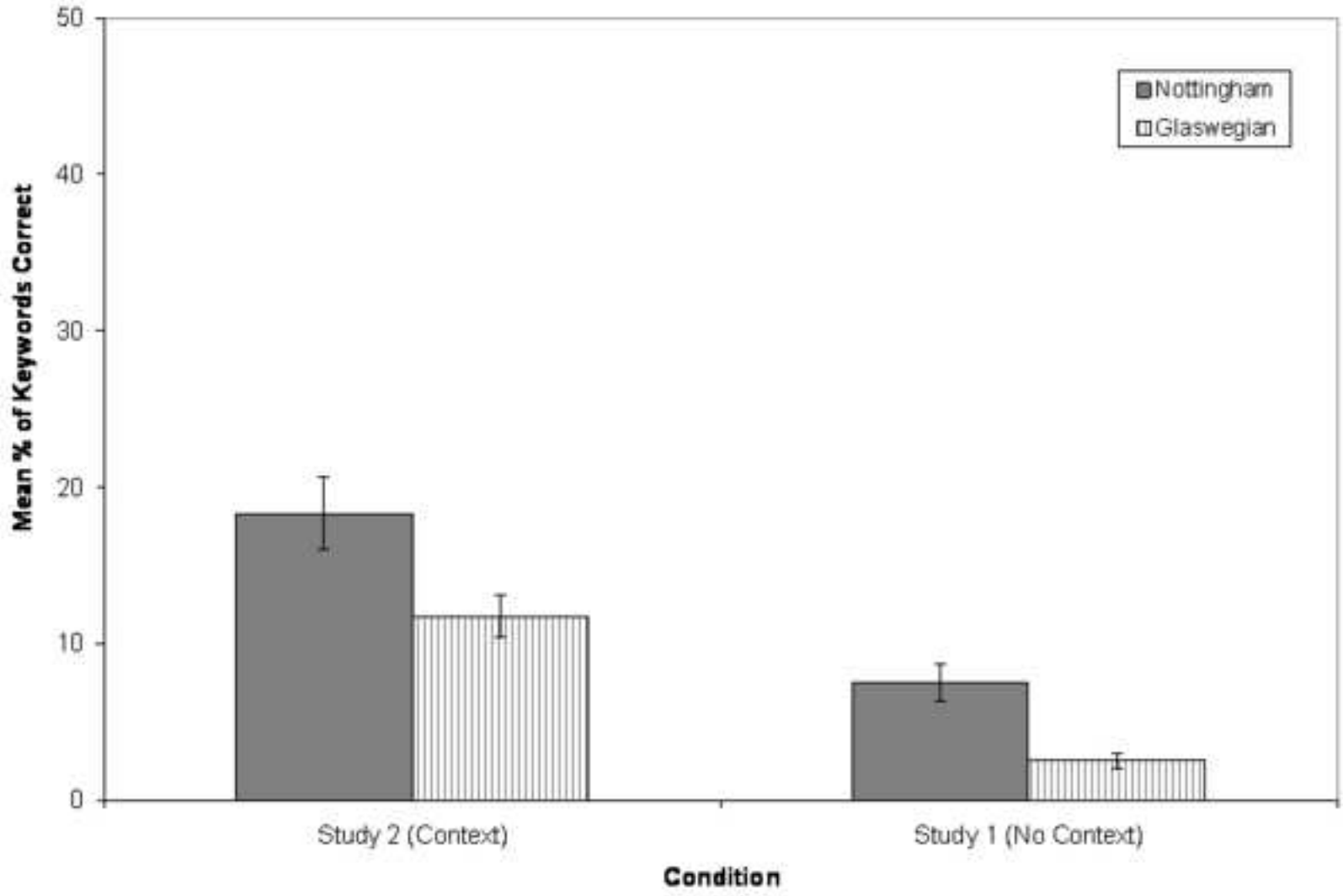
ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT



ACCEPTED MANUSCRIPT