**Sub-regional ‘Other Accent’ Effects on Lay Listeners’ Speaker Identification Abilities:
a voice line-up study with speakers and listeners from the North East of England**

Accepted manuscript for publication in the International Journal of Speech, Language and the Law published by Equinox Publishing. It will appear in a revised form subsequent to editorial input by Equinox Publishing.

<https://journals.equinoxpub.com/index.php/IJSLL/article/view/37340>

Copyright: Equinox Publishing Ltd.

**Abstract**

Previous studies have shown that listeners perform worse in speaker identification experiments when they are unfamiliar with the accents of the speakers. Such effects have been documented for listeners hearing unfamiliar foreign languages (language familiarity effect) and unfamiliar regional accents (‘other accent’ effect). The present study investigates the ‘other accent’ effect at a sub-regional level. Listeners from three different localities (Newcastle, Sunderland and Middlesbrough) within the same greater dialectal region (the North East of England) participated in one of three target-present voice line-ups using samples spoken by speakers from one of the three localities. Listeners who heard a voice line-up in their own local accent (in-group listeners) missed the target speaker’s voice significantly less often than listeners who heard a voice line-up comprised of speakers of one of the other two local accents (out-group listeners). The proportions of correct hits and false alarms were approximately similar across in-group and out-group listeners.

*Keywords:* ‘other accent’ effect, in-group bias, voice line-up, voice parade, North East England

**Introduction**

Previous psychological research has shown that members of certain social groups tend to perceive individual members of the same group as being more easily distinguishable than members of other social groups, across a range of domains. This phenomenon has become known as the ‘out-group homogeneity effect’ or – for other-race faces – as the ‘other race’ effect (Chance & Goldstein 1981; Park & Rothbart 1982; Judd & Park 1988). Myers (2001) describes a similar effect with regard to hearing unfamiliar accented voices as the ‘all sound alike’ phenomenon. With regard to speaker identification, the so-called ‘language familiarity effect’ shows that listeners’ speaker identification performance can be significantly diminished when speakers talk in an unfamiliar language versus a language that the listeners are familiar with (see Fecher & Johnson 2018). In line with this, there is evidence that listeners perform less well at distinguishing between speakers of unfamiliar accents than between speakers of accents with which the listener is familiar. This has become known as the ‘other accent’ effect (see Stevenage & Neil 2014). It must be mentioned, however, that the distinction between familiar and unfamiliar accents is gradual rather than clear-cut and categorical. The Oxford English Dictionary (2018) defines the word *familiar* as ‘well known from long or close association’ and the word *unfamiliar* as ‘not having knowledge or experience of’. Hence, an unfamiliar accent is, broadly speaking, an accent to which a listener has had no (significant) exposure. A familiar accent, on the other hand, is an accent to which the listener has been frequently exposed, e.g. one which the listener has grown up around and may also use him-/herself. Further reasons for a higher degree of familiarity with a particular accent are, for example, close contact with speakers of this accent or frequent exposure to this accent via the media (see Stevenage, Clarke & McNeill 2012).

**Previous research on the ‘other accent’ effect**

Several studies have demonstrated the ‘other accent’ effect and its influence on listeners’ speaker identification abilities. Vanags, Carroll & Perfect (2005) tested the ability of native speakers of Australian English to recognise speakers with either Australian English or British English accents. Listeners performed with significantly greater accuracy when identifying speakers of Australian English than when identifying speakers of British English. In another study, Kerstholt, Hansen, van Amelsvoort & Broeders(2006) tested the speaker identification abilities of Dutch listeners using samples of speech from either Standard Dutch speakers or Dutch speakers with a strong regional accent from The Hague. The listeners in Kerstholt et al. (2006) came from areas of the Netherlands other than The Hague. Although the effect was only marginally significant, listeners’ speaker identification abilities were found to have been more accurate when hearing the Standard Dutch target speaker than the target speaker with the unfamiliar regional accent. Stevenage et al. (2012) tested the ability of listeners from Glasgow and Southampton to identify Scottish speakers with Glaswegian accents and Southern English speakers with Standard Southern British English (SSBE) accents. Both listener groups performed with better accuracy when identifying a target voice from a line-up composed of speakers of their own accent, but Southern English listeners had greater difficulty in identifying the target when hearing the Scottish set of speakers than Scottish listeners had when presented with the Southern English set of speakers. The authors considered different amounts of exposure to the ‘other accent’ as a likely explanation for the observed asymmetry effect[[1]](#endnote-1) as, via the media, Scottish listeners are assumed to have more exposure to SSBE accents than Southern English listeners have to Glaswegian accents. Atkinson (2015) investigated UK listeners’ accent classification abilities (i.e. their abilities to recognise and label regional accents from different parts of the UK) in the form of an online experiment. Listeners from the North East of England were found to be significantly better at recognising North Eastern English accents than were listeners from other parts of the UK. Another group of listeners, who were not from the North East but were familiar with North Eastern English accents, performed less accurately than the North Eastern cohort but more accurately than listeners unfamiliar with North East accents. When listeners had to classify British English accents other than those from the North East of England, all groups performed similarly well. At a sub-regional level, listeners from the areas of Newcastle (Tyneside), Sunderland (Wearside) and Middlesbrough (Teesside) all showed significantly better accent recognition abilities for their own local accent than for the other local accents. Among listeners from Tyneside, however, their superiority at the task was limited to only one of the two Tyneside voices presented. Atkinson (2015) also carried out speaker identification experiments with speakers and listeners from the North East of England. In contrast to the present study, Atkinson used three mixed-voice line-ups which all contained speech samples of speakers of different local North Eastern English accents. Two of the three voice line-ups (one containing a target speaker from Tyneside and one with a target speaker from Wearside) were target-present line-ups. The voice line-up with a Teesside target speaker was a target-absent line-up in which listeners had to recognise that the target speaker’s voice was missing. All voice line-ups included eight voice samples of different speakers. While hearing the voice line-up containing the target speaker from Wearside (which also included three distractor speakers from Tyneside, two distractor speakers from Teesside, and two distractor speakers from Wearside), listeners from the North East of England hardly ever mistook a distractor speaker with a different local accent for the target speaker. Three of the five non-matching distractor speakers were never mistaken for the target speaker by listeners from the North East and the remaining two distractor speakers were only mistaken for the target in a single instance. The fact that so many distractor voices in this line-up were never (or only once) mistaken for the target speaker after 41 listeners had completed the task shows that the nominal size of the line-up (= total number of voices in the line-up, here: eight) was considerably larger than the effective size of the line-up (= the number of speakers that listeners considered for their decision). Listeners from the North East of England could with relative ease eliminate the distractor speakers whose local accents did not match the target speaker’s. This could explain the unusually high correct identification rate of 60 percent by listeners from the North East of England in Atkinson’s voice line-up. At a sub-regional level, Atkinson did not find any significant differences with regard to North Eastern English listeners’ speaker identification abilities, though non-significant trends in all three line-ups showed listeners who shared the target speaker’s local accent to have performed somewhat better than listeners who did not. As the lack of any significant sub-regional ‘other accent’ effects on listeners’ speaker identification abilities in Atkinson’s study could be due to the mixed-voice line-up design that was used, the present study investigated North Eastern English listeners’ speaker identification abilities with separate voice line-ups for the local accents of Newcastle (Tyneside), Sunderland (Wearside), and Middlesbrough (Teesside).

**Sub-regional speaker identification experiment**

In order to test whether ‘other accent’ effects on lay listeners’ speaker identification abilities exist at a sub-regional level, a voice line-up experiment with speakers and listeners from the same greater dialectal region (the North East of England) was undertaken. Participants were either from Newcastle, Sunderland, or Middlesbrough (localities within the North East of England, see Figure 1).

**Figure 1:** Map of Great Britain with a close-up of the North East of England. Newcastle (Tyneside) is in the northern zone, Sunderland (Wearside) in the central zone, and Middlesbrough (Teesside) in the southern zone of the North East.



*Why the North East?*

The North East of England is one of the better described regions in Britain in terms of phonological variation and can be demarcated as a linguistic region with relative ease[[2]](#endnote-2). Even though production studies carried out in the major urban centres of the North East show considerable variation across the region (Kerswill 1987, Watt 1998, Llamas 2001, Burbano-Elizondo 2008, Atkinson 2011), people from outside the area often perceive accents from the North East as a single variety (see Beal, Burbano-Elizondo & Llamas 2012, Montgomery 2007). The speaker identification experiment presented in this article is part of a larger project entitled ‘The Use and Utility of Localised Speech Forms in Determining Identity: Forensic and Sociophonetic perspectives’ (TUULS, supported by ESRC Grant ES/M010783/1). For this project, 120 participants (40 each from Newcastle, Sunderland, and Middlesbrough) were recorded in a range of different contexts (such as reading a short story, participating in a sociolinguistic interview, and taking part in a mock police interview). All speakers were broadly termed ‘working class’ and belonged to one of two age groups (18-25 or 40-65). The sample was balanced for sex and ‘routinised mobility’ defined as the average (non-cumulative) distance which participants travelled in a typical week. Ten speakers from each locality (30 in total) were recorded a second time at a later date during which they also took part in a series of perception tests, with the voice line-up experiment described in the following section being one such test.

*Speech samples*

All of the speech samples used in the experimental line-up were collected from 18-25-year-old males drawn from the TUULS corpus of North Eastern English. From these, three separate voice line-ups of eight speakers were constructed – one composed of speakers from Newcastle, one of speakers from Sunderland, and another of speakers from Middlesbrough. For each of the three localities, there were recordings of at least ten male speakers available. Since the present voice line-up experiment is part of a larger project (TUULS) and recordings of eight speakers per locality were required for the voice line-ups, the choice of speakers to select was limited. Care was taken not to include any speakers who sounded markedly different from the others (e.g., due to having a very low-pitched voice or atypical manner of speaking) in these line-ups.

*Material*

Three different eight-person voice line-ups were constructed following the McFarlane guidelines for the creation of voice line-ups as closely as possible (see Home Office 2003, Nolan 2003). The voice samples in each line-up came from speakers with a local accent from either Newcastle, Sunderland, or Middlesbrough. Voice samples had a duration of 60 seconds each and consisted of spontaneous speech taken from mock police interview recordings from the TUULS corpus. The spontaneous speech samples used for listener familiarisation with a particular target speaker’s voice each had a duration of 30 seconds and were taken from sociolinguistic interviews from the TUULS corpus. Importantly, only sections containing non-identifying information were taken from each interview (i.e. the wording did not reveal any information about where within the North East speakers were from). All recordings were made with a ZOOM H4n digital recorder (settings: 44.1 kHz, 16 bit) and a DPA 4088 directional headset microphone, with all samples normalised to -3 dB.

*Listeners*

30 listeners from the North East of England (ten each from Newcastle, Sunderland, and Middlesbrough) participated in the voice line-up experiment. Listeners were either aged between 18-25 or 40-65, with none reporting any hearing difficulties. The set was roughly balanced for sex, age group, and level of routinised mobility (based on the average (non-cumulative) distance travelled in a typical week). All listeners participated in a single voice line-up and, when asked, stated that they had not recognised any of the speakers in the line-up as someone they had already heard before the experiment.

*Pretest*

A pretest with nine (later uninvolved) listeners was carried out in order to ensure that none of the selected voices stood out from the others in the respective line-ups for any reason. Pretest listeners followed the same procedure as in the main experiment. In the line-up containing Sunderland-accented voices, two pretest listeners correctly identified the target speaker’s voice. The pretest listeners who heard the Newcastle and Middlesbrough-accented line-ups, however, all failed to do so. Consequently, the target speaker in the Sunderland line-up was replaced by another target speaker and the line-up was re-tested with different pretest listeners, yielding similar results as the pretest of the other two line-ups. The pretest procedure was also used to try to balance the difficulty of the three voice line-ups. Note that as the described voice line-up experiment was part of a larger project, there were constraints on how many pretest listeners could be tested.

*Allocation of listeners to voice line-ups*

Each listener participated in one of the three voice line-ups. The voices in each line-up either matched with the listener’s own local accent (in-group condition) or consisted of voice samples from eight speakers who spoke with one of the other two local accents (out-group condition). The pseudo-random allocation of listeners to voice line-ups is illustrated in Table 1.

**Table 1:** Allocation of the 30 listeners to the three voice line-ups. Shaded areas show how many listeners heard a voice line-up in their own local accent (in-group condition).

|  |  |
| --- | --- |
|  | Listeners from |
|  | *Newcastle* | *Sunderland* | *Middlesbrough* |
| *Newcastle* voice line-up | **4** | 3 | 3 |
| *Sunderland* voice line-up | 3 | **4** | 3 |
| *Middlesbrough* voice line-up | 3 | 3 | **4** |

*Familiarisation with the voice of the target speaker*

All listeners took part in the experiment in single, separate sessions. During the familiarisation task, participants listened to a 30-second recording of a speaker who had to be recognised later (but listeners were not informed about the purpose of the experiment at this stage). Each listener listened to one recording of one target speaker (who was either from Newcastle, Sunderland, or Middlesbrough). This task was described to the listeners as a ‘calibration’ procedure. They were instructed to listen carefully to the recording, to pay attention to the voice, and to feed back whether the recording quality and volume level were satisfactory. Participants listened to the recordings on Sennheiser HD 280 Pro closed-cup headphones. After the familiarisation task, listeners took part in an unrelated recording session which involved reading aloud a list of words and a short story, retelling the short story in their own words, and describing six different picture stories. This acted as a separator task to avoid potential short-term memory influences on the results of the speaker identification test.

*Voice line-up presentation*

After participating in the unrelated recording tasks, listeners were informed of the purpose of the speaker identification experiment. Each listener was then presented with a voice line-up composed of speakers of the same local accent as the target speaker that the listener had heard in the familiarisation task. Per locality, there were five different versions of the same line-up, varying only in the pseudo-randomised presentation order of the speech samples (in order to counteract possible position effects). Each presentation order was heard by two listeners in the main experiment. Although all voice line-ups were target-present line-ups (in which a voice sample of the target speaker was included), listeners were informed that the target speaker heard earlier might or might not be present in the line-up. Voice line-ups were programmed and presented in OpenSesame 3.2, a program which has its own scripting language but also allows for input in Python (Mathôt, Schreij & Theeuwes 2012). After the screen with the instructions, listeners saw eight consecutive screens containing one button each (labelled from A to H). By clicking the buttons, participants were played 30-second-long samples of each speaker in the line-up. After listening to all eight speakers in the line-up, listeners were given the opportunity to listen to some or all of the speakers again as often as they wished, with these speech samples each lasting a full minute (rather than just 30 seconds). The decision to limit the duration of the speech samples to 30 seconds in the first run-through was taken after multiple pretest listeners had pointed out that the experiment felt too long, introducing an increased risk of fatigue effects.

During the voice line-up experiment, listeners wore Sennheiser HD280 Pro headphones and were allowed to take notes. When ready to give an answer, they moved on to the next screen where they could either select one of the speakers (A-H), state that they believed that the target speaker had not been present in the line-up, or indicate that they could not answer this question. After choosing an option, listeners were presented with a screen which asked them to rate their level of confidence in the decision on a visual analogue scale (with 100 possible values), with the extreme ends of the scale labelled as ‘not sure at all’ and ‘very sure’. Finally, listeners were presented with a screen allowing them to provide additional comments about why they had picked or discounted particular speakers. Finally, participants were asked to indicate whether they knew any of the stimuli speakers heard prior to the experiment, with all participants responding that they had not.

**Listeners’ familiarity with different local accents in the North East of England**

After taking part in the speaker identification experiment, listeners rated their familiarity with the local accents of Newcastle, Sunderland, and Middlesbrough on five-point Likert scales, with the first point representing ‘not at all familiar’ and the last (i.e. point 5) ‘very familiar’.

**Table 2:** Listeners’ familiarity with the local accent in Newcastle, Sunderland and Middlesbrough. Mean scores (and standard deviations). Shaded areas show listeners’ familiarity ratings regarding their own local accent.

|  |  |  |  |
| --- | --- | --- | --- |
| **Listeners from ↓** | *Newcastle accent* | *Sunderland accent* | *Middlesbrough accent* |
| **Newcastle** | **4.8** (0.4) | **3.7** (0.9) | **2.9** (1.1) |
| **Sunderland** | **3.7** (1.2) | **4.4** (1.1) | **2.5** (0.8) |
| **Middlesbrough** | **3.5** (0.8) | **2.0** (0.5) | **4.9** (0.3) |

The familiarity ratings in Table 2 show that listeners in each locality rated their own local accent as the accent they were most familiar with, although listeners from Sunderland rated lower on average and were more variable in how they rated their own local accent than listeners from the other locations. Since Newcastle is the biggest city in the North East of England and the socioeconomic centre of the region as a whole (see Beal et al. 2012), it may be expected that listeners from across the North East would be quite familiar with the Newcastle accent, and this is indeed reflected in the familiarity ratings. Finally, it is noteworthy that listeners from Newcastle and Sunderland generally indicated that they were reasonably familiar with each other’s accents. One likely explanation for this is a greater amount of contact between residents of these cities, given their closer geographic proximity, than with residents of Middlesbrough. Newcastle and Sunderland are only 10 miles (17 km) apart from each other whereas Middlesbrough is 31 miles (50 km) south of Newcastle and 23 miles (38 km) south of Sunderland (see Figure 1).

**Results**

Data analyses were carried out in R (version 3.5.1) and with MedCalc’s Comparison of Proportions Calculator, which performs N-1 Chi-squared Tests, as recommended by Campbell (2007) and Richardson (2011). This modified Chi-squared Test can deal with small sample sizes and has more statistical power than other tests for this type of data. All voice line-ups in the present study were target-present line-ups, giving three potential outcomes in the speaker identification experiment: a *correct hit* when the listener picked the target speaker’s voice, a *false alarm* when the listener picked a distractor speaker’s voice, and a *miss* when the listener did not pick any voice in the line-up.

*Performance of in-group vs out-group listeners*

The proportions of correct hits were not statistically different between in-group listeners, who listened to a line-up in their own local accent, and out-group listeners, who listened to a line-up in a local accent that was not their own (N-1 Chi-squared Test, two-tailed, χ2= 2.148, df = 1, p = 0.1427). The same was true of the proportions of false alarms (N-1 Chi-squared Test, two-tailed, χ2= 0.537, df = 1, p = 0.4637). In-group listeners, however, missed the target speaker’s voice significantly less often than did out-group listeners (N-1 Chi-squared Test, two-tailed, χ2= 4.322, df = 1, p = 0.0376). The differences in percent between the performance of in-group and out-group listeners are shown in Table 3.

**Table 3:** Percentages of (correct) hits, misses and false alarms by in-group and out-group listeners.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **(correct) hits** | **misses** | **false alarms** |
| **in-group**  |  33.33% |  8.33% |  58.33% |
| **out-group** |  11.11% | 44.44% | 44.44% |
| **difference** | not significant | significant | not significant |

With regard to listeners’ decision criteria trends, Table 3 shows that in-group listeners were more accepting. They achieved more correct hits but also made more false alarms. Out-group listeners were more conservative in their decisions, yielding fewer false alarms but also fewer correct hits.

*Influence of routinised mobility, sex and age*

Regarding proportions of correct hits, no significant differences were found between mobile speakers, who travelled a minimum distance of ten miles from home on a weekly basis, and non-mobile listeners, who tended to stay within their own localities (N-1 Chi-squared Test, two-tailed, χ2= 0.032, df = 1, p = 0.8572; mobile listeners: 18.75% correct hits, non-mobile listeners: 21.43% correct hits). Male and female listeners performed similarly well (N-1 Chi-squared Test, two-tailed, χ2= 0.032, df = 1, p = 0.8572; male listeners: 21.43% correct hits, female listeners: 18.75% correct hits). There was also no significant difference in speaker identification performance between listeners in the younger and older age groups (N-1 Chi-squared Test, two-tailed, χ2= 1.165, df = 1, p = 0.2804; younger listeners: 28.57% correct hits, older listeners: 12.5% correct hits). No significant differences were found in the proportions of false alarms and misses between groups based on age, sex, or mobility. Unfortunately, groups could not be broken down any further for analyses, due to small sample sizes.

**Table 4:** Correct hits, false alarms and misses (in percent) of mobile vs non-mobile listeners, male vs female listeners and younger vs older listeners. Total number of listeners = 30 (split in different ways).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **listeners:** | **mobile** | **non-mobile** | **male** | **female** | **younger** | **older** |
| **proportions**  | **16/30** | **14/30** | **14/30** | **16/30** | **14/30** | **16/30** |
| *correct hits* | 18.75 % | 21.43 % | 21.43 % | 18.75 % | 28.57 % | 12.50 % |
| *false alarms* | 62.50 % | 35.71 % | 42.86 % | 56.25 % | 50.00 % | 50.00 % |
| *misses* | 18.75 % | 42.86 % | 35.71 % | 25.00 % | 21.43 % | 37.50 % |

*Listeners’ choices in the Middlesbrough line-up*

Figure 2 illustrates the choices listeners made in the voice line-up with the Middlesbrough-accented voices. Two of the ten listeners in this condition correctly selected the target speaker’s voice, six listeners selected the voice of a distractor speaker, and two listeners believed that the target speaker’s voice was not present in the line-up.

**Figure 2:** Listeners’ choices of particular speakers or the option ‘none’ in the voice line-up with Middlesbrough-accented voices. The numbers on the x-axis represent speaker numbers in the TUULS corpus. Selection frequency of the target speaker (number 4) = diagonally hatched bar. Selection frequencies of particular distractor speakers or the option ‘none’ (i.e. not picking any of the speakers in the line-up) = solid black bars.

*Listeners’ reasons for (not) picking a speaker in the Middlesbrough line-up*

The following ten quotes provide some information about listeners’ reasons for their choice in the Middlesbrough line-up:

‘deeper voice, slower speech, less pronounced’
(in-group listener’s comment; correct hit, speaker 4)

‘had similar age-sounding voice. Familiar accent and same type of description.’
(in-group listener’s comment; correct hit, speaker 4)

‘is the closest but speaker might not have been included’

(in-group listener’s comment; false alarm, speaker 13)

‘Because he did not use many word fillers, such as erm or hmm, and when I heard the original speaker, he did not use that many either. Also, I thought the original speaker sounded local to my accent, so I chose the one that sounded most like what I hear day-to-day around my area.’

(Sunderland listener’s comment; false alarm, speaker 13)

‘picked this one because it sounded the same as before’

(in-group listener’s comment; false alarm, speaker 102)

‘it was between F and no one. I thought F [speaker 102] had a similar accent to the person speaking in the first clip’

(Newcastle listener’s comment; false alarm, speaker 102)

‘they do not have the accent or vocal mannerism of the speaker from the beginning’

(Sunderland listener’s comment; miss (none))

‘a lot of them sounded much the same to me’

(Newcastle listener’s comment; miss (none))

‘the accents towards the end were less regional and definitely didn’t belong to the opening speaker. If the opening speaker was part of this selection, he was towards the beginning and my feeling is about seventy per cent that he was speaker B [speaker 28]’ [The target speaker’s voice was in the third last position in this voice line-up].

(Newcastle listener’s comment; false alarm, speaker 28)

‘I picked this speaker because it resonated with some thoughts when I heard the speaker the first time, such as wondering where this speaker was from and the fact that this speaker sounded quite bold and sure of himself’ (Sunderland listener’s comment; false alarm, speaker 111)

*Listeners’ choices in the Sunderland line-up*

Listeners’ choices in the line-up with the Sunderland-accented voices are illustrated in Figure 3. Four of the ten listeners in this condition correctly selected the target speaker’s voice, four listeners selected the voice of a distractor speaker, and two listeners believed that the target speaker’s voice was not present in the line-up.

**Figure 3:** Listeners’ choices of particular speakers or the option ‘none’ in the voice line-up with Sunderland-accented voices. The numbers on the x-axis represent speaker numbers in the TUULS corpus. Selection frequency of the target speaker (number 122) = diagonally hatched bar. Selection frequencies of particular distractor speakers or the option ‘none’ (i.e. not picking any of the speakers in the line-up) = solid black bars.

*Listeners’ reasons for (not) picking a speaker in the Sunderland line-up*

The following ten quotes provide some information about listeners’ reasons for their choice in the Sunderland line-up:

‘the speaker’s voice was distinctly Geordie, while the others (except for A [speaker 119]) didn’t appear so’

(in-group listener’s comment; correct hit, speaker 122)

‘more of a Sunderland accent’

(in-group listener’s comment; correct hit, speaker 122)

‘I have picked this speaker because he is the only one that sounds like he has a Middlesbrough accent. At the beginning, I was sure I was listening to a Middlesbrough accent’

(Middlesbrough listener’s comment; correct hit, speaker 122)

‘sounds more familiar’

(Newcastle listener’s comment; correct hit, speaker 122)

‘he sounded the most familiar. I felt he spoke at the same speed as the voice I heard earlier and with about the same degree of confidence’

(in-group listener’s comment; false alarm, speaker 96)

‘the voice I picked had soft tones, not quite Geordie but stronger than a Teesside accent so therefore I came to the conclusion that the accent was the same as the first and that it was from the Washington area’

(Middlesbrough listener’s comment; false alarm, speaker 96)

‘sounds like both recordings are from somewhere South of Tyne, like Wearside or maybe Teesside, like Middlesbrough’

(Newcastle listener’s comment; false alarm, speaker 96)

‘didn’t recognise the voice from earlier’

(Middlesbrough listener’s comment; miss (none))

‘the voices I’ve heard here all sounded younger than the original and they didn’t sound as broad’

(Newcastle listener’s comment; miss (none))

‘B [speaker 34] or D [speaker 96] possible, difficult to establish which speaker as not expected to recall that information from initial speaker although feels like the nearest match’

(in-group listener’s comment; false alarm, speaker 34)

*Listeners’ choices in the Newcastle line-up*

Figure 4 depicts the choices listeners made in the line-up with the Newcastle-accented voices. Five of the ten listeners in this condition selected the voice of a distractor speaker, four listeners did not select any of the voices because they believed that the target speaker’s voice was not present in the line-up and one listener did not select any of the voices because he could not remember the target speaker’s voice. In this voice line-up, none of the ten listeners correctly selected the voice sample of the target speaker. (Potential reasons for why this might have happened will be discussed later).

**Figure 4:** Listeners’ choices of particular speakers or the option ‘none’ (incl. no answer, see section below) in the voice line-up with Newcastle-accented voices. The numbers on the x-axis represent speaker numbers in the TUULS corpus. Target speaker = speaker 144. Selection frequencies of particular distractor speakers or the option ‘none’ (i.e. not picking any of the speakers in the line-up) = solid black bars.

*Listeners’ reasons for (not) picking a speaker in the Newcastle line-up*

The following ten quotes provide some information about listeners’ reasons for their choice in the Newcastle line-up:

‘the speakers’ accents were not as strong as the speaker at the beginning although I was not sure to begin with if speaker A [speaker 135] was the speaker at the beginning’

(in-group listener’s comment; miss (none))

‘the more I listened to the other speakers, the less I could recall the original speaker. I also was looking for what the original speaker had said but didn’t hear the same words.’

(Middlesbrough listener’s comment; miss (none))

‘in the original recording, I felt as though I remember listening to slang. In the later recordings, it was not present. The only doubts I had were with recordings H [speaker 135] and D [speaker 108]’

(Sunderland listener’s comment; miss (none))

‘none of them sounded exactly the same, even though they were all of a similar age and from a similar area’ (Sunderland listener’s comment; miss (none))

‘can’t remember the first speaker’

(Sunderland listener’s comment; miss (no answer: the listener was not sure enough to provide an answer))

‘C [speaker 135] and H [speaker 142] very similar, others sounded different’

(in-group listener’s comment; false alarm, speaker 135)

‘other voices sounded sort of less aggressive in their accent’

(Middlesbrough listener’s comment; false alarm, speaker 135)

‘accent was too light whereas beginning speaker had a rougher accent’

(Middlesbrough listener’s comment; false alarm, speaker 124)

N/A

(in-group listener’s comment; false alarm, speaker 142)

‘some of the speakers weren’t as clear as the original. They were breathier and the pitch of some speakers was too high as the original was quite low in pitch’

(in-group listener’s comment; false alarm, speaker 108)

*Performance of listeners from Newcastle, Sunderland and Middlesbrough*

Figure 5 shows that in all three localities, some listeners returned correct hits, some missed the target speaker’s voice and some made false alarms. Therefore, all possible outcomes in a target-present speaker identification experiment were observed in all three listener groups.

**Figure 5:** Performance of listeners from Newcastle, Sunderland and Middlesbrough. (Correct) hits (selection of the target speaker’s voice) = diagonal hatching, misses (no selection of any voice) = solid black, false alarms (selection of a distractor speaker’s voice = grey.

*Listeners’ confidence in their decisions*

**Figure 6:** Listeners’ confidence in the decisions they made in the voice line-up experiment. Confidence in answers that turned out to be correct vs confidence in answers that turned out to be false.

As can be seen in Figure 6, the difference in listener confidence ratings associated with correct vs false answers in the Sunderland voice line-up is not statistically significant (Wilcoxon rank sum test, W = 7, p = 0.3524). Nor is the difference in confidence ratings associated with correct vs false answers in the Middlesbrough voice line-up (Wilcoxon rank sum test, W = 6, p = 0.7111). For the Newcastle line-up, this difference could not be assessed, as none of the ten participants who listened to this line-up managed to correctly identify the target speaker.

*Participants’ listening times per speaker in the voice line-up*

Tables 5-7 show average times in seconds that participants spent listening to the speakers in the three voice line-ups. In the Middlesbrough line-up (Table 5) and the Sunderland line-up (Table 6), listeners who heard speakers in their (the listeners’) own local accent listened the longest on average to the target speaker’s voice, whereas listeners who heard a line-up in a different local accent spent the longest time on average listening to a distractor speaker’s voice. Such a difference between in-group and out-group listeners was not found in the line-up with the Newcastle-accented voices. However, it will be recalled that this is the line-up in which none of the ten listeners gave a correct answer.

**Table 5:** Average time in seconds that participants who listened to a same-accent line-up (in-group) and participants who listened to a voice line-up in another local accent (out-group) listened to the eight speakers in the Middlesbrough line-up. The longest average listening time per group is printed in bold and the target speaker’s column is shaded in grey. Standard deviations are shown in parentheses.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| MBR line-up | *Spk 4**target* | *Spk 6* | *Spk 11* | *Spk 13* | *Spk 28* | *Spk 102* | *Spk 105* | *Spk 111* |
| in-group | **45.2** (21.6) | 30.0 (0.0) | 31.1 (2.2) | 38.8 (17.7) | 40.4 (14.3) | 31.1 (2.1) | 30.0(0.0) | 31.8 (2.4) |
| out-group | 31.9(3.0) | 33.2 (4.0) | **41.0** (26.3) | 34.2(7.0) | 31.1 (2.0) | 38.6(8.8) | 31.9(3.2) | 33.0(3.9) |

**Table 6:** Average time in seconds that participants who listened to a same-accent line-up (in-group) and participants who listened to a voice line-up in another local accent (out-group) listened to the eight speakers in the Sunderland line-up. The longest average listening time per group is printed in bold and the target speaker’s column is shaded in grey. Standard deviations are shown in parentheses.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUN line-up | *Spk 122**target* | *Spk 24* | *Spk 34* | *Spk 47* | *Spk 96* | *Spk 112*  | *Spk 119*  | *Spk 123* |
| in-group | **54.4** (28.5) | 30.0 (0.0) | 33.1(6.2) | 30.0(0.0) | 40.3(16.1) | 30.0(0.0) | 30.0(0.0) | 30.0(0.0) |
| out-group | 40.3(13.7) | 30.4(1.1) | 30.3(0.8) | 30.7(1.7) | 40.2(21.8) | **41.3**(27.3) | 30.5(1.2) | 30.3(0.8) |

**Table 7:** Average time in seconds that participants who listened to a same-accent line-up (in-group) and participants who listened to a voice line-up in another local accent (out-group) listened to the eight speakers in the Newcastle line-up. The longest average listening time per group is printed in bold and the target speaker’s column is shaded in grey. Standard deviations are shown in parentheses.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| NCL line-up | *Spk 144**target* | *Spk 108* | *Spk 124* | *Spk 130* | *Spk 135* | *Spk 140* | *Spk 142* | *Spk 145* |
| in-group | 31.4 (2.8) | 40.8(21.7) | 30.0(0.0) | 30.0(0.0) | **55.4**(31.0) | 30.0(0.0) | 47.8(30.3) | 30.7(1.4) |
| out-group | 33.7(5.5) | 35.2(7.4) | 33.0(3.6) | 30.0(0.0) | **37.1**(7.4) | 31.6(3.0) | 31.6(3.8) | 31.3(2.5) |

**Discussion**

This study investigated the existence of ‘other accent’ effects on listeners’ speaker identification abilities at a sub-regional level. Listeners from three different localities (Newcastle, Sunderland, and Middlesbrough) within the same greater dialectal region (the North East of England) participated in one of three voice line-ups. There was one line-up per locality, each composed of eight 18-25 year-old male speakers. Four listeners from each of the three localities listened to a voice line-up in their own local accent, while the other six listeners (three and three) in each locality listened to a voice line-up in a different local accent. The results show that participants who listened to a voice line-up in their own local accent (in-group condition) missed the target speaker’s voice significantly less often than did participants who listened to a voice line-up in one of the other two local accents (out-group condition). In comparison, correct hit rates and false alarm rates were more similar across groups. As will be recalled from the Section ‘*Previous research on the “other accent” effect’,* Kerstholt et al. (2006) found that listeners who had to identify a target speaker in a voice line-up containing Standard Dutch voices marginally outperformed listeners who had to identify a target speaker in a voice line-up with a strong unfamiliar regional accent from The Hague. A closer look at the results of the target-present line-ups revealed that the percentage of correct hits was lower and the percentage of false alarms was higher in the unfamiliar regional accent condition compared to the Standard Dutch condition. In the target-absent line-ups, listeners made more false alarms after listening to samples of regionally-accented speech from The Hague versus Standard Dutch. Stevenage et al. (2012), who examined ‘other accent’ effects in voice line-up experiments with Glaswegian and Southern British listeners and speakers, reported higher false alarm rates among listeners presented with unfamiliar accents. In Atkinson (2015), listeners from the North East of England correctly recognised speakers with North Eastern English accents at significantly higher rates than did listeners from other parts of the UK. Atkinson did not find significant differences in listeners’ speaker identification performance on a sub-regional level; however, his study used a mixed-voice line-up design (as described above).

Although there are some differences between the findings of the aforementioned studies and the findings of the present study with regard to what kind of error (i.e., miss or false alarm) was made more frequently when listeners tried to identify a speaker with an unfamiliar (sub-regional) accent, studies overall agree that listeners’ speaker identification performance tends to suffer when the target speaker speaks with an unfamiliar accent or a different local accent. Nevertheless, a direct comparison of the results of the aforementioned studies is difficult, as different experimental settings were used. For example, studies differed in the type of memory they tested (intentional vs incidental memory) and the length of the retention interval between the first exposure to the target speaker’s voice and the presentation of the voice line-up. In a real forensic case in the Netherlands, earwitnesses were shown to be able to identify an offender’s voice up to a year after the incident had taken place, with the accuracy of the result later backed up by DNA evidence (Vermeulen 2018). Listeners’ memory, however, generally deteriorates in the time between the first exposure to a voice that has to be recognised later and the presentation of a voice line-up. The process of forgetting is complex and can depend not only on the listener but also on the characteristics of the speaker’s voice (see Kerstholt et al. 2006).

Unfortunately, the potential asymmetry effect (i.e. whether North East listeners in general perform fairly well in the Newcastle-accented line-up because of listeners’ proportionately higher familiarity with the local accent of the largest city in their region) cannot be analysed further, as none of the participants who listened to the Newcastle line-up were able to correctly identify the target speaker. More data are needed in order to determine whether asymmetry effects (as described in Stevenage et al. 2012) occur at a sub-regional level. With regard to listeners’ accent recognition abilities, however, Atkinson (2015) provided some evidence for such an effect. While listeners from the areas of Newcastle (Tyneside), Sunderland (Wearside), and Middlesbrough (Teesside) all showed significantly better accent recognition abilities for their own local accent on average, the superiority of Tyneside listeners over other North East listeners was limited to one of the two presented Tyneside voices. This could be because listeners from the North East of England generally have a reasonable level of familiarity with the local accent of the largest city in the region (see Atkinson 2015).

In the present study, a closer look at the Newcastle line-up revealed that half of the participants in this condition did not choose any voice as a match for the target speaker, with four of those listeners stating that the target speakers’ voice had not been present and another stating that he could not remember the target speaker’s voice. In both the Middlesbrough and Sunderland line-ups, only two listeners failed to pick any of the presented voices as a match for the target speaker (with all four listeners stating that the target speaker’s voice had not been present in the line-up). When the fairness of the Newcastle line-up was assessed in the form of a pretest, both listeners who tested the Newcastle line-up did pick a voice when asked to identify the target speaker. One reason for the lack of correct answers in the Newcastle line-up could be a higher amount of intra-speaker variability in the speech of the Newcastle target speaker (i.e. his voice sounded different in the voice sample used for the familiarisation phase and the voice sample used in the line-up, for example, due to a style shift). Evidence for the latter is reflected in three of five listeners’ comments on why they did not pick a speaker in this line-up. One listener said that the accents in the voice line-up were ‘not as strong’ as the accent of the target speaker, another listener pointed out that none of the voices in the line-up sounded ‘exactly the same’ as the target speaker, and a third listener felt that the target speaker had used more ‘slang’ than the speakers in the voice line-up. Additionally, a listener who had picked a distractor speaker in this line-up stated that the target speaker at the beginning had a ‘rougher accent’.

Listeners’ comments also revealed some evidence for Myers’ (2001) ‘all sound alike’ phenomenon at a sub-regional level: a listener from Newcastle who listened to the Middlesbrough line-up commented on the voices that ‘a lot of them sounded much the same to me’.

Listeners’ feedback on the Sunderland line-up also proved to be interesting, with three listeners who correctly identified the target speaker stating contradicting reasons for their choice. One listener from Sunderland picked the target speaker’s voice because it sounded ‘distinctly Geordie’ (i.e. from Newcastle), another listener from Sunderland reported that the target speaker had ‘more of a Sunderland accent’, and a listener from Middlesbrough said he had picked the target speaker ‘because he is the only one that sounds like he has a Middlesbrough accent. At the beginning, I was sure I was listening to a Middlesbrough accent’.

Other listeners who gave incorrect answers in the speaker identification experiment were able to place the local accent of the target speaker to some extent. A listener from Middlesbrough who had participated in the Sunderland line-up stated that the target speaker had an accent that was ‘not quite Geordie but stronger than a Teesside accent […] and that it was from the Washington area’. Sunderland (Wearside) indeed lies between Newcastle (Tyneside) and Middlesbrough (Teesside) and the target speaker actually had links to Washington (Wearside), as his grandparents had moved there. Distractor speaker 96, however, whom the listener from Middlesbrough had mistaken for the target speaker, had not reported having any links to the Washington area. A listener from Newcastle, who had also picked distractor speaker 96 from the Sunderland line-up, commented that it ‘sounds like both recordings are from somewhere South of Tyne like Wearside or maybe Teesside like Middlesbrough’. Finally, a listener from Sunderland that participated in the Middlesbrough line-up failed to identify the target speaker and misclassified the local accent, commenting that ‘[…] I thought the original speaker sounded local to my accent, so I chose the one that sounded most like what I hear day-to-day around my area’. Listener feedback shows that the link between listeners’ accent classification abilities and their abilities to identify speakers does not always straightforwardly point in the same direction. Thus, more research on the link is needed.

Tables 5 and 6 show that in-group listeners hearing the Middlesbrough and Sunderland line-ups on average listened to the target speaker for longer than any distractor speakers, while out-group listeners in both line-ups on average listened longer to one of the distractor speakers. This could indicate that in-group listeners (but not out-group listeners) ‘short-listed’ the target speaker’s voice more frequently, giving it greater consideration in their final decision. Table 7 shows that in-group and out-group listeners in the Newcastle line-up listened to one of the distractor speakers for greater durations, on average, than they did to the target speaker. No listeners, however, were able to give a correct answer in this line-up.

The results regarding the listeners’ confidence levels in their decisions versus their accuracy in the speaker identification task show that listeners’ confidence is a poor predictor of accuracy. Although not statistically significant, the trend pointed even in the opposite direction, in that listeners were slightly more confident when their answer was actually incorrect. The finding that confidence is not a good predictor of accuracy in speaker identification experiments is in line with previous studies (Orchard & Yarmey 1995; Öhman, Eriksson & Granhag 2010; Atkinson 2015; Braun 2016) and it has clear implications for casework: we should be careful not to let listeners’ confidence about their decisions influence our levels of confidence in their judgements.

**Conclusion**

The reported sub-regional ‘other accent’ effects should be kept in mind when voice line-ups are carried out in forensic cases where the victim or witness has a different (local) accent from that of the voices in the line-up. Care should be taken, however, not to impose too many further constraints on the *construction* of voice line-ups, as these could render the practical use of voice line-ups in forensic cases impractical, not to say impossible. Under current guidelines (see e.g. McFarlane guidelines[[3]](#endnote-3)), it can already be difficult to source enough suitable recordings of potential distractor speakers.

More research is needed to investigate sub-regional ‘other accent’ effects on a larger scale and to further examine the link between listeners’ speaker identification abilities and their abilities to classify accents. It is possible that further acoustic measurements may reveal why particular voices in the line-ups were or were not mistaken for the target speaker.

**Limitations**

A cautious interpretation of the results of this study is necessary as its sample size was limited by practical constraints. Furthermore, only target-present voice line-ups were carried out and the retention interval between listeners’ familiarisation with the target speaker and the presentation of the voice line-up was rather short (around 10 minutes per participant). Listeners, however, were informed that the target speaker might or might not be present in the line-up and unrelated distractor tasks were used to counteract short-term memory effects.

**Acknowledgements**

This work was supported by ESRC Grant ES/M010783/1. We thank all participants and pretest listeners for taking part in our study. Furthermore, we thank our anonymous reviewers for their helpful feedback on this article.

**References**

Atkinson, J. 2011. *Linguistic variation and change in a North-East border town: a sociolinguistic study of Darlington.* PhD thesis. Newcastle University. <https://theses.ncl.ac.uk/dspace/bitstream/10443/1214/1/Atkinsonj11.pdf>

Atkinson, N. 2015. *Variable factors affecting voice identification in forensic contexts.* PhD thesis. University of York. <http://etheses.whiterose.ac.uk/13013/>

Beal, J. C., Burbano-Elizondo, L., Llamas, C. 2012. *Urban North-Eastern English. Tyneside to Teesside.* Edinburgh, Edinburgh University Press.

Braun, A. 2016. *The speaker identification ability of blind and sighted listeners. An empirical investigation.* (Doctoral thesis, Philipps-Universität Marburg, Germany, 2015). Wiesbaden, Springer VS. <https://link.springer.com/book/10.1007/978-3-658-15198-0>

Burbano-Elizondo, L. 2008. *Language variation and identity in Sunderland.* PhD thesis, University of Sheffield. <http://etheses.whiterose.ac.uk/3644/>

Campbell, I. 2007. Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations. *Statistics in Medicine*, 26, 3661-3675. <https://doi.org/10.1002/sim.2832>

Chance, J. E. & Goldstein, A. G. 1981. Depth of processing in response to own and other race faces. Journal of Personality and Social Psychology, **7** (3), 475-480. <https://doi.org/10.1177/014616728173017>

Fecher, N. & Johnson, E. K. 2018. Effects of language experience and task demands on talker recognition by children and adults. *Journal of the Acoustical Society of America,* 143 (4), 2409-2418. <https://doi.org/10.1121/1.5032199>

Home Office. 2003. Advice on the use of voice identification parades. *UK Home Office Circular 057/2003 from the Crime Reduction and Community Safety Group, Police Leadership and Power Unit.* [http://webarchive.nationalarchives.gov.uk/20130308000037/ http://www.homeoffice.gov.uk/about-us/corporate-publications-strategy/home-office-circulars/circulars-2003/057-2003/](http://webarchive.nationalarchives.gov.uk/20130308000037/http%3A/www.homeoffice.gov.uk/about-us/corporate-publications-strategy/home-office-circulars/circulars-2003/057-2003/)

Judd, C. M. & Park, B. 1988. Out-group homogeneity: judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology*, 54 (5), 778-788. <http://dx.doi.org/10.1037/0022-3514.54.5.778>

Kerstholt, J. H., Hansen, N. J. M., van Amelsvoort A. G. & Broeders, A. P. A. 2006. Earwitnesses: effects of accent, retention and telephone. *Applied Cognitive Psychology,* 20 (2), 187-197. <https://doi.org/10.1002/acp.1175>

Kerswill, P. 1987. Levels of linguistic variation in Durham. *Journal of Linguistics,* 23, 25-49.

<https://doi.org/10.1017/S0022226700011026>

Llamas, C. 2001. *Language variation and innovation in Teesside English.* PhD Dissertation. University of Leeds. <http://etheses.whiterose.ac.uk/2939/>

Mathôt, S., Schreij, D. & Theeuwes, J. 2012. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavior Research Methods,* 44 (2), 314-324. <https://doi.org/10.3758/s13428-011-0168-7>

Montgomery, C. 2007. *Northern English dialects: a perceptual approach.* PhD thesis, University of Sheffield. <http://etheses.whiterose.ac.uk/1203/>

Myers, D. G. 2001. *Psychology.* New York, NY. Worth Publishers.

Nolan, F. 2003. A recent voice parade. *The International Journal of Speech, Language and the Law,* 10 (2), 277-291. <http://dx.doi.org/10.1558/sll.2003.10.2.277>

Öhman, L., Eriksson, A. & Granhag, P. A. 2010. Mobile phone quality vs. direct quality: how the presentation format affects earwitness identification accuracy. *The European Journal of Psychology Applied to Legal Context,* 2 (2), 161-182. <https://doaj.org/article/7c4d2abe0c2b48e7a08499819d1ae15d>

Orchard, T. L. & Yarmey, A. D. 1995. The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology,* 9 (3), 249-260. <https://doi.org/10.1002/acp.2350090306>

Oxford English Dictionary 2018. Oxford University Press. Retrieved from <https://en.oxforddictionaries.com/> on 30 September 2018.

Park, B. & Rothbart, M. 1982. Perception of out-group homogeneity and level of social categorization. Journal of Personality and Social Psychology, 42 (6), 1051-1068. <http://dx.doi.org/10.1037/0022-3514.42.6.1051>

Richardson, J. T. E. 2011. The analysis of 2 x 2 contingency tables – Yet again. *Statistics in Medicine,* 30, 890. <https://doi.org/10.1002/sim.4116>

Stevenage, S. V., Clarke, G. & McNeill, A. 2012. The ‘other-accent’ effect in voice recognition. *Journal of Cognitive Psychology,* 24 (6), 647-653. <https://doi.org/10.1080/20445911.2012.675321>

Stevenage, S. V. & Neil, G. J. 2014. Hearing faces and seeing voices: the integration and interaction of face and voice processing. *Psychologica Belgica,* 2014, 54, 266-281. <http://doi.org/10.5334/pb.ar>

Vanags, T., Carroll, M. & Perfect, T. J. 2005. Verbal overshadowing: a sound theory in voice recognition? *Applied Cognitive Psychology,* 19, 1127-1144. <https://doi.org/10.1002/acp.1160>

Vermeulen, J. 2018. Case Report. Constructing a voice line-up with multiple ear-witnesses. *Paper presented at the 27th conference of the International Association for Forensic Phonetic and Acoustics.* Huddersfield, UK.

Walker, P. M. & Tanaka, J. W. 2003. An encoding advantage for own-race versus other-race faces. *Perception,* 32 (9), 1117-1125. <https://doi.org/10.1068/p5098>

Watt, D. 1998. *Variation and change in the vowel system of Tyneside English.* PhD dissertation. University of Newcastle-upon-Tyne. <https://theses.ncl.ac.uk/dspace/bitstream/10443/350/1/Watt98.pdf>

1. Asymmetry effects have also been observed in the visual domain with regard to the ‘other race’ effect. Some studies show that the difference between people’s relatively poorer ability to recognise other-race faces and their ability to recognise own-race faces is smaller in people who have more interracial experience (see Walker & Tanaka 2003). [↑](#endnote-ref-1)
2. Although linguists are generally happy to talk about the North East as a relatively easy-to-define linguistic region, it can be difficult to say what all the NE varieties in fact have in common, and which phonetic/phonological features distinguish them from neighbouring varieties. [↑](#endnote-ref-2)
3. In the present study, we followed the McFarlane guidelines as closely as possible, but we were unable to follow them exactly. For example, we did not have speech recordings available from 20 young male speakers per locality from which we could select suitable distractor speakers. What ultimately matters is the alignment of the voices in the line-up in terms of accent and other parameters, not demographics/metadata of the speakers themselves. [↑](#endnote-ref-3)