The processing of unfamiliar accents in a competing talker task

Brianne Senior

Department of Linguistics
University of British Columbia
LING 449 Honours Essay

Supervisor: Dr. Molly Babel

April 2017

Abstract

Listeners' ability to pay attention to one speaker against a background of other speech – a phenomenon dubbed the cocktail party problem – varies according to properties of the speech streams and the listener. Although a number of acoustic and experiential factors that contribute to a listener's ability to successfully segregate two simultaneous speech signals have been identified, there are competing predictions about the role unfamiliar accents may play in this process. To this end, familiar Canadian-accented voices and unfamiliar British-accented voices were used in a competing talker task using the coordinate response measure. Listeners heard two different talkers simultaneously read sentences in the form of "[command] [colour] [preposition] [letter] [number] [adverb]" (e.g., "Lay blue at C4 now") and had to report the coordinate from the talker who said blue. Results from Canadian listeners indicate that on all but the most challenging trials, listeners did best when attending to an unfamiliar British-accented target against a familiarly-accented masker, but do not do similarly well when forced to ignore this unfamiliar accent. These results suggest listeners can easily tune out a familiar accent but are unable to do the same with an unfamiliar accent.

1. Introduction

Conversations do not always occur under the most ideal listening conditions. Often, the target speech is uttered amongst a background of other sounds, including other speech. In such "multi-talker" situations, the listener must first divide their attention amongst the competing speech signals in order to identify which signal contains the critical information. Then, they must allocate their attentional resources to the target signal while "tuning out" irrelevant, masking speech. A classic real-world example of the importance of this selective attention is clearly seen in the realm of aviation. Pilots and other aviation professionals often find themselves confronted with situations in which they must attend to multiple, simultaneous signals and determine which of these signals are personally relevant (Ericson, Brungart, & Simpson, 2004). For example, a commercial pilot may need to monitor signals from both air traffic control and in-flight cabin crew, while a military pilot may need to mediate directives from other aircraft and command centres. Ambient cockpit noise may further obscure the incoming signals, adding to the complexity of this task.

In a seminal paper, Cherry (1953) described the challenge a listener faces in understanding one individual's speech while other speakers are simultaneously speaking as the "cocktail party problem." He found that when a listener was forced to disentangle a mixture of two concurrent spoken passages, the listener could successfully divorce the two messages but only with great difficulty; in his original study, the listener had to relisten to the messages upwards of ten times in order to separate and reassemble the original passages. Much of the challenge in recognizing what a talker is saying under speech-in-speech listening conditions arises from the fact that competing speech signals often produce both informational and energetic masking (Brungart, 2001b). Energetic masking involves two acoustic signals with the same frequencies occurring at the same point in time; in this way, one or more portions of the acoustic signals are made physically imperceptible. With informational masking, however, a listener can clearly detect all parts of the competing signals, but they may be unable to understand, segment, and make sense of them. Whereas energetic masking predominates in speech-in-noise situations and the effects of this masking can easily be overcome by increasing the signal-to-noise ratio, the role of signal-to-noise ratios in speech-in-speech situations is more complex and may have a more minimal impact on intelligibility (Brungart, 2001b).

Researchers have utilized many different methods to investigate how listeners solve the cocktail party problem. Cherry himself used two different methods to probe this question. In one experiment, Cherry (1953) presented the same mixed, two-talker audio track to both of his listener's ears, and the listener then had to reconstruct the individual messages. He also used a method that has come to be known as the dichotic listening task. In this task, a listener hears one message in their left ear and simultaneously hears a different method in their right ear; they are then told to repeat (shadow) one of the messages and are later questioned about the speech in the unattended (unshadowed) ear. One of the major problems with the dichotic listening task is that the amount of information garnered from each participant is minimal compared to the amount of time expended.

The coordinate response measure or CRM (Moore, 1981; see also Bolia, Nelson, Ericson, & Simpson, 2000 for a more recent CRM corpus) is an alternative method for exploring speech perception in speech-in-noise and speech-in-speech situations. In the CRM task, the listener is exposed to two simultaneous speech signals in the form of "Ready [call sign], go to [colour] [number] now," mirroring commands a military pilot might receive. One of the speakers reads the target call sign, and it is this talker that the listener must attend to. Then, the listener must accurately segment and track the target talker in order to respond with the correct letter-number coordinate. The CRM has recently been used to examine speech perception in multi-talker situations. It has been used to establish the relative contributions of energetic and informational masking (Brungart, 2001b; Brungart, 2005), spatial location (Ericson & McKinley, 1997), voice familiarity (Johnsrude et al., 2013), listener native language (Cooke, Garcia Lecumberri, & Barker, 2008), and contralateral ear masking (Brungart, Simpson, Darwin, Arbogast, & Kidd Jr, 2005) to speech perception. It has also been used in speech perception tasks with more than two competing talkers (Brungart, Simpson, Ericson, & Scott, 2001). Unlike the dichotic listening task which almost exclusively involves selective attention, the CRM task involves both divided and selective attention. In the first part of the task, the listener must split their attention between both signals to find the target signal. Then, they must selectively attend to only the target signal and segment the target talker's utterance from that of the masking talker. The CRM overcomes the slow rate of data collection problem faced by the dichotic listening task as several CRM trials can be run in quick succession. Additionally, by examining the coordinate responses, the CRM allows for a detailed analysis of error type. For example, did the listener track the wrong speech stream, respond with information from both speech streams, or respond with information that was present in neither signal? There is also evidence that the CRM is comparable to other speech intelligibility assessments such as the articulation index (Brungart, 2001a), a measure that predicts intelligibility based on the acoustic structure of the signal.

Yet, the CRM is not without its flaws. Like the dichotic listening task, it fundamentally lacks ecological validity outside of aviation situations. Whereas the dichotic listening task suffers primarily because a listener is not normally talking while listening to their conversational partner speak, the CRM's drawback is that its scripted sentences do not resemble natural conversation. To overcome these challenges, a more naturalistic "simulated cocktail party" has been proposed in which multiple talkers each tell a story and the listener has to answer questions about the story, the answers to which come from various speakers (Hafter, Xia, & Kalluri, 2013). However, this method is relatively new and requires a somewhat complex lab setup in which multiple loudspeakers are arranged spatially around the listener to simulate different talkers. Ultimately, all of these methods seek to answer how a listener divides their attention amongst two or more talkers in order to identify and focus on the most important and relevant information.

There are a number of factors that may influence listener performance in a competing talker or speech segregation task (e.g., see Bronkhorst, 2000 for a detailed review). Much research has been done on the acoustic factors inherent to the competing speech signals, termed "signal-driven" factors. Signal-driven factors include properties of both the target (talker to be attended to) and masker (one or more background voices that should be ignored). In general, the more distinct the vocal characteristics of the two signals, the easier it is to separate the competing signals; speech segregation is hardest when both the target and masker speech originate from the same talker (Brungart, 2005; Cooke et al., 2008). Differences in fundamental frequency between the two signals seem to provide some release from masking, as performance is highest when the target and masker are different genders (Brungart, 2005; Brungart et al., 2001). Similarly, differences in the relative intensity of the target and masker can also aid a listener in this task (Bronkhorst, 2005; Brungart, 2001b; Brungart et al., 2001). When both target and masker are the same amplitude, there are no level differences for listeners to exploit as a cue to segregation. That said, not all level differences are equally effective; given two concurrent talkers at different intensities, Brungart (2001b) found it was more challenging to pay attention to the quieter speaker than it is to attend to the louder.

The perceived spatial location of the two signals also plays an important role in speech segmentation. Separating the apparent location of the sound source in space can improve performance in this task, at least when only two talkers are involved (Ericson & McKinley, 1997). Ericson and McKinley (1997) had listeners complete a speech-in-speech intelligibility task in either diotic (same two-talker signal presented to both ears; no spatial separation of sound sources), dichotic (one talker in each ear; sound sources separated by 90 degrees azimuth), or directional (same two-talker signal presented to both ears but talkers seemed to be originating from different directions in space on the horizontal plane) listening conditions. They found that listeners were best able to separate competing talker signals when they were presented dichotically (90 degrees azimuth separation) and that even a spatial separation of 45 degrees azimuth improved performance compared to diotic presentation. Similar results were obtained by Peng, Zhang, and Wang (2012); their results indicated that speech-in-speech identification was consistently better when the target and masker signals were presented dichotically rather than diotically, regardless of the number of competing talkers. Yet, it is also possible for certain aspects of the acoustic signal to make it more difficult for the listener to process in the presence of competing speech. For example, acoustically "degrading" the target speech signal (e.g., by using synthesized speech) makes it more difficult to attend to (Sinatra, Sims, Najle, & Bailey, 2012).

Research has also begun to look at the role of knowledge-driven factors in speech segmentation. Knowledge-driven factors are those that are intrinsic to the listener rather than the acoustic signal, and they include aspects of the listener's language experience that may assist them in segmenting multiple simultaneous speech streams. For example, listeners are better at attending to speech if it is in their native language (Cooke et al., 2008). They perform especially well if the target speech is in their native language and the masking speech is in an unfamiliar language; Garcia Lecumberri & Cooke (2006) found that English sentences embedded in a background of competing Spanish voices were more intelligible to native English-speaking listeners than when the competing voices were also in English. Conversely, listeners who spoke both Spanish and English performed equivalently in these two scenarios. Knowledge-driven factors also seem to operate on an individual voice level. Johnsrude et al. (2013) found that familiarity with a talker's voice improved performance in a speech segregation task; listeners could both selectively attend to and ignore a familiar voice, but could not do the same for an unfamiliar voice.

Listeners also do better when they have some knowledge of what the target or masker signals will resemble. Brungart et al. (2001) found that when listeners were given details about the vocal characteristics (i.e., gender) of the target speaker prior to the task, they were able to use this prior knowledge to aid themselves in selectively attending to the talker. This knowledge does not have to be made explicit to the listeners. Cherry (1953) discovered that the difficulty in separating two concurrent spoken passages was considerably lessened when the passages were made highly predictable (e.g., by using idiomatic expressions such as "beating around the bush" or "on the brink of ruin"), suggesting the listener was able to use their knowledge of which words tend to occur together to make the task easier.

Most of the work in the area of multi-talker situations has looked at the effects of signal-driven and knowledge-driven factors separately. Accents provide an avenue to explore the interface between signal-driven and knowledge-driven processes. Foreign-accented speech involves phonetic and prosodic deviations from familiar, local-accented speech. It has been shown to be more difficult to understand and to require greater processing time (Munro & Derwing, 1995). In noisy listening conditions, the intelligibility of foreign-accented speech seems to particularly suffer. Rogers, Dalby, and Nishi (2004) showed that even highly proficient non-native English speakers were still less intelligible than native English speakers. Similar results have been found by Gordon-Salant, Yeni-Komshian, Fitzgibbons, Cohen, and Waldroup (2013), who found that segregation cues (e.g., talker-masker gender differences) normally used in multi-talker situations were less effective when the target spoke with a moderate Spanish accent. It has been suggested that processing foreign-accented speech requires extra attention and effort (Rogers et al., 2004; Van Engen & Peelle, 2014). These processing difficulties are not limited to non-native accents. Although the effects are not as great as those for non-native accented speech, unfamiliar native accents do take longer to process in noise than familiar, local accents (Adank, Evans, Stuart-Smith, & Scotti, 2009), and there seems to exist a minimal "processing cost" for unfamiliar regional accents even in quiet listening conditions (Floccia, Goslin, Girard, & Konopczynski, 2006). Given these results, it is unsurprising that Van Engen and Peelle (2014) argue accents may form a kind of "acoustically degraded" speech.

If foreign- or unfamiliarly-accented speech is a form of "degraded" speech, it follows that it should be more difficult to attend to in a competing talker situation. If a listener is unfamiliar with a particular accent, it could also potentially influence performance. As Cooke et al. (2008) argue, "A listener whose knowledge of English is restricted to one specific accent may be less able to assign speech sounds from other accents to the target or the background source" (pp. 425-426). Taken together, these findings suggest that it should be harder for a listener to attend to an unfamiliar accent. Yet, research on the role of accents in competing talker situations offers conflicting results. Sinatra (2012) examined the role of British-accented and American-accented English speech in a dichotic listening task. They found that there was no difference in the amount of unattended information reported when the voice in the unattended channel spoke British-accented or American-accented English. Moreover, the variation offered by a foreign accent may serve as an additional cue to segregation and actually facilitate performance in a multi-talker situation. In hypothesizing how listeners may solve the cocktail party problem, Cherry (1953) listed "differing accents" as a potential contributing factor. This hypothesis seems to be supported by Joshi, Iyer, and Gupta (2013). They found that listeners are better at identifying a speaker with a distinct accent amongst a background of talkers speaking with a different but shared accent.

So, do the systematic variations in foreign-accented speech truly render it more difficult to process in challenging, speech-in-speech listening conditions? Or, alternatively, does the variation unfamiliarly-accented speech offers make it more salient and therefore easier to attend to? I test these competing hypotheses in a speech segregation task with two competing talkers speaking either Canadian-accented English (familiar accent) or British-accented English (unfamiliar). British-accented English is likely unfamiliar to most students at the University of British Columbia. Of the more than 12,000 international students on campus, just little over 300 of them are from the United Kingdom (UBC PAIR, 2015). The phonetic differences between Canadian-accented and British-accented English include both segmental and prosodic differences. For example, one major distinguishing characteristic of Canadian-accented English is in the PRICE and MOUTH vowels. In Canadian English, these starting points for these vowels are usually raised before voiceless consonants so that they are pronounced with a more mid central vowel like [əɪ] and [əʊ], respectively (Wells, 1982). By contrast, speakers of most British accents – barring those from the northernmost parts of England – pronounce these vowels with low starting points. Another major phonetic difference between these two dialects is that Canadian English is known for merging the vowels in PALM, CLOTH, THOUGHT, LOT, and START so that they are all pronounced with a low back unrounded vowel which stands in contrast to the low back rounded vowels in many varieties of British English (Wells, 1982). Another distinct characteristic of Canadian-accented English is the so-called "Canadian shift" in which many of its front lax vowels are becoming lowered and increasingly retracted (Clarke, Elms, Youssef, & 1995). There has also been some evidence that the rising and falling pitch contours may be particularly steep in British-accented English when compared to American-accented English (Yan & Vaseghi, 2002), and it seems likely these differences would also extend to Canadian English.

In this experiment, I seek to add to our knowledge of how listeners solve the cocktail party problem by examining the contribution of familiar and unfamiliar accents. I compare performance across four different target-masker conditions – Canadian-Canadian, Canadian-British, British-Canadian, and British-British – in a competing talker task using the CRM. I predict that listeners unfamiliar with British-accented English will find these accents harder to attend to against a background of familiarly-accented speech. If this is true, listeners should be less accurate in reporting target coordinates when they are asked to attend to a British-accented target masked by familiar Canadian-accented speech than when asked to attend to a Canadian-accented target masked by British-accented speech. Three different target-to-masker intensity ratios (TMRs) and talkers of both genders are included in order to examine the role of target-masker accents with respect to these known factors in speech segregation.

2. Methods

2.1 Stimuli

2.1.1 British-accented English speakers

Eight (four male, four female) speakers of British English were taken from the Grid corpus (Cooke et al., 2006). The Grid corpus consists of 34 British English talkers each reading 1000 different sentences in the form of "[command] [colour] [preposition] [letter] [number] [adverb]." For example, a typical sentence from the Grid corpus would be "Place blue with C 9 now." Within these sentences, the colour, letter, and number are keywords, whereas the command word, preposition, and adverb are fillers. The Grid corpus sentences are a variation on the typical CRM sentences originally used by Moore (1981). While the original CRM sentences were in the form of "Ready [call sign], go to [colour] [number] now," the sentences in the Grid corpus offer greater phonetic variation (see Table 1). This offers a couple of advantages. The greater phonetic variation, particularly in the fillers, makes the task slightly more naturalistic and helps avoid acoustic “echoes” that can be created by two of the same fillers being spoken at the same time.

Table 1

Sentence structure for the Grid corpus. Keywords are identified with asterisks.

command	colour*	preposition	letter*	digit*	adverb
bin	blue	at	A-Z excluding W	1-9, zero	again
lay	green	by			now
place	red	in			please
set	white	with			soon

Note. From Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*(5), p.2422

Although all talkers in the Grid corpus are reported as being native English speakers with a “range of English accents,” they speak primarily northern varieties of English, as they were recruited from Sheffield, England (J. Barker, personal communication, August 22, 2016). The authors report that all talkers spent the majority of their lives in England, with three obvious exceptions: two talkers were from Scotland and one from Jamaica. These three talkers were readily identified by their accents and thus were not selected for use in this experiment. Talkers with hyperarticulated speech or excessively creaky voice quality were also not selected. From the remaining voices, two phonetically-trained listeners selected four male and four female talkers (speakers 2, 6, 11, 13, 16, 22, and 23 from the Grid corpus). These talkers were selected for their relative homogeneity with respect to FACE, GOAT, and STRUT vowels,[1] as well as the quality

[1] While there are other vowels that distinguish various British accents (e.g., the BATH vowel), these vowels were always reduced given the structure of the carrier sentences and could not be compared across talkers.

of their recordings and relative pitches.

This subset of eight Grid corpus talkers left a pool of 8000 possible sentence. Sentences with the command word “bin” were ignored, as this is a term likely unfamiliar to Canadian English listeners. Additionally, I removed all sentences with the letter “z”, as Canadian English often varies between “zed” and “zee” pronunciations due to influence from American English (Zeller, 1993). Likewise, all sentences with “H” were not used, due to the British alternation between “haytch” and “aytch,” the former of which is not generally used in Canadian English. All sentences with the number “0” were also removed, as this could be pronounced either “zero” or “oh.”

From this subset of sentences, I randomly selected 100 sentences per British English talker using the RANDOM function in MS Excel. These 800 sentences were trimmed in Praat (Boersma & Weenink, 2016) so that the individual files contained no extra silence before or after the sentence, and were peak amplitude normalized. Any sentences with errors (e.g., lengthy hesitations, disfluencies, missing words, excessive noise or recording bumps, entire sentence not in the recording window) were discarded.

2.1.2 Canadian-accented English speakers

Taking into account the aforementioned modifications, I generated a list of all possible keywords and fillers. From this master list, I created eight individual lists of 150 randomly selected sentences. Each list was mutually exclusive (i.e., sentences in one list were not included in any other list).

I recorded eight speakers of Canadian English, four female and four male. All were self-identified native speakers of Canadian English, had spent all or the vast majority of their life in Canada, and the only variety of English they spoke was Canadian-accented English. All but one of the speakers were born in Canada; the other moved to Canada at six months of age. None of the speakers reported any speech, hearing, or language disorders. The mean speaker age was 29 years old (range of 18-48, nearly the same range as in Grid corpus).

Each Canadian English speaker was assigned their own list of 150 sentences. Recording was done in E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) using a headmounted microphone with a SoundDevices USB PreAMP in a sound-attenuated cubicle. Speakers first heard eight sample sentences to familiarize them with the task, and none of these sentences were included in any of the aforementioned lists. The sample sentences varied from 1.53 - 1.87 seconds long, roughly corresponding to the middle two-thirds of all file durations in the Grid corpus. The sample sentences included examples of each colour and filler word.

Speakers completed 150 randomly-ordered trials corresponding to their individual sentence list. On each trial, speakers saw a sentence on the screen which was displayed for 3000 ms. They were asked to read each sentence aloud as naturally as possible, while fitting within the three second recording window. After each sentence, they were given the opportunity to rerecord the sentence or to proceed to the next sentence. Speakers were instructed at the beginning of the experiment to only rerecord sentences for which they made mistakes or did not produce within the three second recording window. Only the most recent recording attempt was recorded. Speakers received a self-paced break every 30 sentences. Following the recording session, they completed a survey about their language background. The entire session took approximately 20-25 minutes, and all speakers participated voluntarily.

Recordings were trimmed so that the onset of the recording corresponded with the onset of the sentence. No interval of silence was assumed in the case of initial voiceless stops. Sound files were downsampled to 25 kHz to match the sampling rate of the files in the Grid corpus and peak amplitude normalized. Sentences with noise, hesitations, errors, or recording bumps were discarded.

Two additional native English speakers were recorded reading 50 sentences each to be used in the practise section. None of these sentences were identical to the ones produced by the other 16 speakers.

2.1.3 Acoustic analysis of speakers

Fundamental frequency data for each of the 16 speakers was estimated using REAPER (https://github.com/google/REAPER). Average fundamental frequency was calculated from a random sample of 30 sentences for each speaker,[2] excluding unvoiced periods. Females (British females: *M* = 184 Hz, *SD* = 11; Canadian females: *M* = 194 Hz, *SD*= 13) had higher fundamental frequencies than males (British males: *M* = 118 Hz, *SD* = 13; Canadian females: *M* = 123 Hz, *SD* = 10), and fundamental frequencies were similar across accents for each gender.

2.2 Pretesting intelligibility: Transcription task

The remaining 1783 sound files (range 88 - 144 sentences per talker, mean of 111 sentences per talker) were pretested in a transcription task to get a baseline measure of intelligibility. Each sentence was embedded in accent-specific speech-shaped noise at 5 dB SNR. This noise was

[2] For expediency, only 30 sentences per speaker were sampled out of the pool of all possible sentences for each speaker.

generated from all the sentences produced by talkers of that particular accent (i.e., 762 British-accented English sentences and 1021 Canadian-accented English sentences). 31 participants completed a transcription task in which they were asked to type out each sentence in standard English orthography. One participant did not finish this task as they found the typing physically uncomfortable. Three participants were removed because English was not their native or dominant language,[3] two participants were removed because they reported speaking British English, and a further two participants were removed for failing to follow the instructions (i.e., they transcribed only the coordinate and not the entire sentence). This left a total of 22 participants. They had a mean age of 20.1 years old (*SD* = 2.5).

Participants were seated in a sound-attenuated cubicle and wore AKG-240 headphones. They were instructed to type the sentences as accurately as possible while excluding punctuation. In order to keep transcriptions consistent, participants were informed that each sentence would contain a letter followed by a number and were asked to press the key that corresponded to the letter or number rather than typing out the entire word (e.g., “c7” rather than “see seven” or “sea seven”). Participants completed 16 blocks, one for each speaker, which appeared in a random order; within each block, participants transcribed 20 sentences randomly selected from the larger pool of all sentences produced by that talker. Participants were given self-paced breaks between each block. The transcription task was presented using E-Prime 2.0 and, on average, took about 45 minutes to complete. Participants were compensated with course credit for their time.

[3] Eligibility requirements for inclusion in the pretest were that listeners reported English as their dominant language, as evaluated by the self-report question, “If English is not your native language, is it your dominant language?” This is different from the criteria for inclusion in the speech segregation task, where listeners had to be not only dominant in English, but also native speakers of English. Using English dominance rather than nativeness as the criterion for the pretest ensured that all sentences were indeed highly intelligible, even to individuals who are not native speakers.

Following the task, spelling errors in transcriptions were corrected and punctuation marks removed to further standardize the transcriptions. I then calculated how many keywords were correctly transcribed in each sentence. Keywords included the colour and letter-number coordinate; coordinates were treated as one unit because in the CRM task, listeners are generally scored on their ability to accurately report the entire coordinate. The proportion of keywords correctly identified was averaged across sentences. Overall sentence intelligibility prior to the exclusion of the low intelligibility sentences was also plotted by talker (see Figure 1). Average intelligibility across all talkers was 88%, with individual talker intelligibility scores ranging from 84-93%. Intelligibility was similar across all talkers of both genders and accents, with Canadian-accented female talkers being the most intelligible ($M = 0.91$, $SD = 0.08$) and Canadian-accented males being the least intelligible ($M = 0.86$, $SD = 0.09$). There was slightly more variability in intelligibility for the British-accented talkers ($SD = 0.11$) than the Canadian-accented talkers ($SD = 0.09$), with listener performance on the British-accented male talkers being the most variable ($SD = 0.11$).

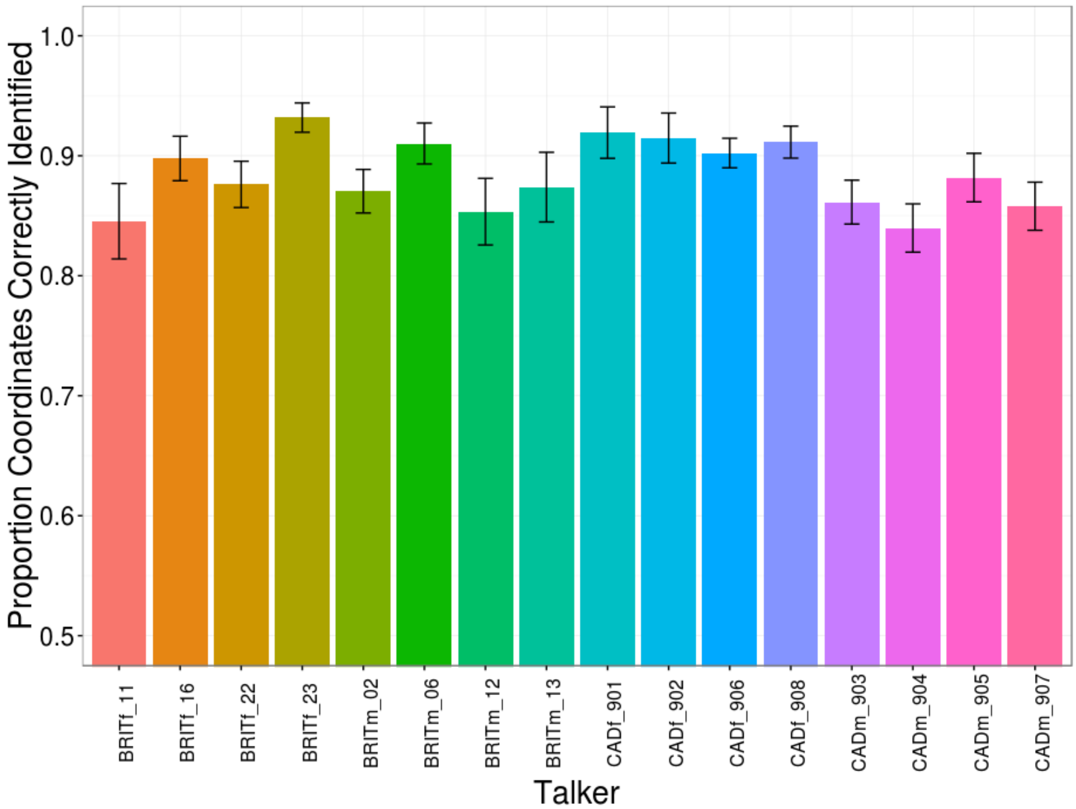

Figure 1. Overall by-talker accuracy prior to removal of the low intelligibility utterances. CAD = Canadian-accented English talker. BRIT = British-accented English talker. f = female. m = male.

Sentences with an average accuracy of less than 90% were discarded. This left 1033 highly intelligible sentences (M = 0.99, SD = 0.03) to be used in the CRM task (see Figure 2).

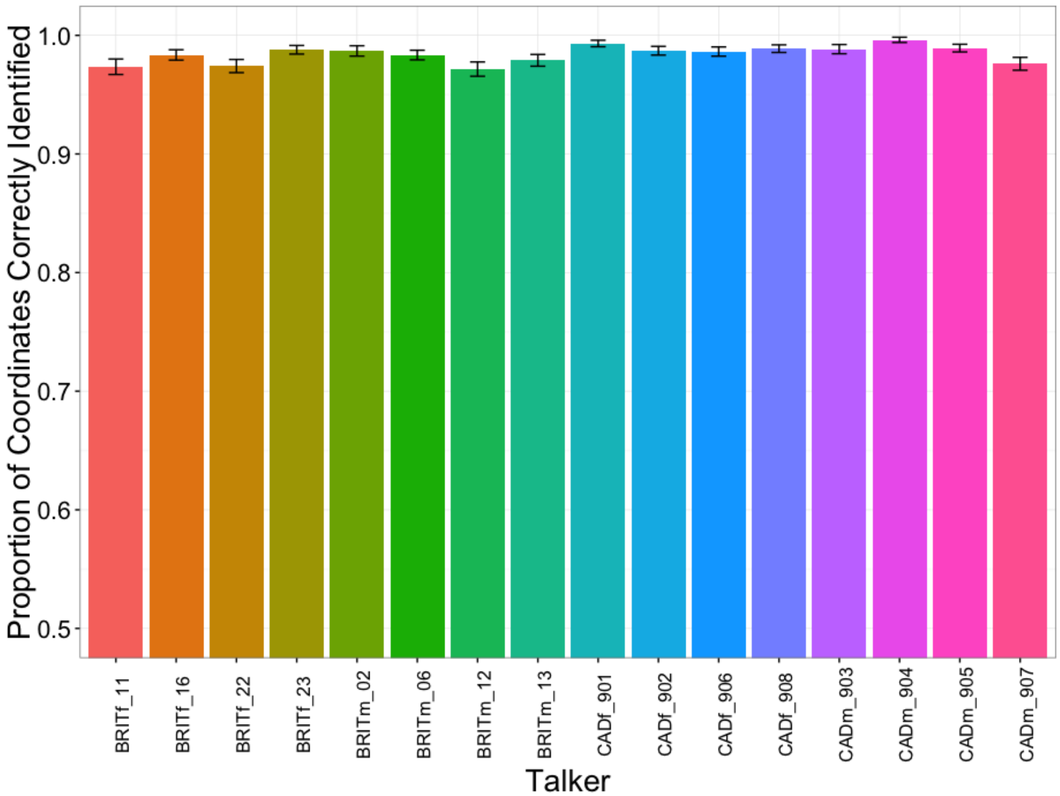

Figure 2. Overall by-talker accuracy after removal of the low intelligibility utterances.

2.3 Multi-talker speech segregation task

The above files (90% accuracy or higher) were then combined to create sentence pairs. Each sound file was matched with a sentence of similar duration so that the two sentences were within 200 ms duration of each other. Within each pairwise combination, the two sentences had different colours and the coordinates were not identical. Across these sentence pairs, there were a mixture of British-British, Canadian-Canadian, and British-Canadian pairings; male-male, female-female, male-female pairings; and +6, -6, and 0 dB TMRs creating 48 different trial types: Target Accent (British, Canadian) x Target Gender (male, female) x Masker Accent

(British, Canadian) x Masker Gender (male, female) x TMR (+6 dB, -6 dB, 0 dB). From these sentence pairings, seven different experimental lists of 240 trials each were created. Each list contained 120 trials with a Canadian-accented talker as the target and 120 trials with a British-accented talker as the target; 120 trials where the target was a female and 120 trials where the target was a male; and 80 trials at each of the three TMRs.

93 participants completed a speech segregation task. 31 were removed because they were not native English speakers. A further four participants were removed because they self-reported speaking British English, and another eight were removed because they were not speakers of Canadian English, as assessed via their answer to the question “Regardless of whether English is your native or dominant language, which variety of English do you speak?” This left 50 participants, 41 of whom were females. Since female listeners greatly outnumbered male listeners, only the data from the 41 female participants was analyzed in order to keep the knowledge of the listener group as homogenous as possible.

The speech segmentation experiment consisted of three parts: (1) practise with no masker, (2) practise with a faint masker, and (3) the main experimental task. In the first practise section, listeners were familiarized with the sentence format. They were informed that each sentence they heard would be in the form of “[command] [colour] [preposition] [letter] [number] [adverb].” They were instructed to type the letter-number coordinate they heard the speaker produce using the keyboard. Listeners completed five practise trials with feedback as to the correct answer. The speaker on all practise trials was not one of the 16 speakers from the main task, and none of the sentences from the main task were used.

In the second practise section, listeners were exposed to an easier version of the full task to familiarize them with the CRM procedure. They heard two simultaneous sentences presented diotically – that is, the same signal was simultaneously presented to both ears – over headphones in which the target-to-masker ratio was +9 dB. They were instructed to type the coordinate from the speaker who said their target colour. Listeners had a maximum of 5000 ms after the offset of the sound file to respond with their coordinate, and they were instructed at the beginning of the experiment to respond even if they were not entirely sure of their answer. Listeners completed 16 practise trials and received feedback as to the correct answer. The talkers in this practise session were different from the 16 talkers in the main task, and none of the sentences in the practise session were used in the main task.

The procedure for the main task was identical to that of the second practise session with the following exceptions: participants did not receive feedback as to the correct response, and the 16 (8 British, 8 Canadian) talkers described above were used. Listeners completed 240 randomly-ordered trials equally distributed amongst the three TMR levels (+6 dB TMR, -6 dB TMR, 0 dB TMR). On all of the trials, the target colour was “blue,” and participants were reminded of this at the beginning of each trial. Participants were given self-paced breaks every 60 trials. Upon completion of the main task, participants completed a short survey about their language background. All parts of the experiment took place in E-Prime 2.0, and the survey was completed electronically through FluidSurveys. The entire experiment took approximately 40 minutes, and participants were compensated with course credit for their time.

3. Results

The focus here is on the less established effects of Talker Accent; thus, in this analysis, the full four-way comparison of which accents are targets and maskers is retained. Talker Gender is included in the analysis, but the coding is simplified to compare same gender and different gender trials, as the effect of talker gender in competing talker situations has been established (e.g., Brungart, 2005; Brungart et al., 2001). Transcription accuracy was averaged over Subject on Accent (Canadian-Canadian, British-British, Canadian-British, and British-Canadian), Gender (same gender, different gender), and TMR. The distribution of transcription accuracy scores was not normal, so these values were converted to rationalized arcsine units (RAU), and RAUs were used as the dependent measure in the analyses below.

The first analysis was an ANOVA examining Accent, Gender, and TMR, with these variables repeated across listeners. There was a main effect of TMR [$F(2, 80) = 246.3$, $p < 0.001$]. Post-hoc Tukey tests (all at $p < 0.001$) established that the +6 dB TMR trials ($M = 95$ RAU, 86% percent, $SD = 23$ RAU) were easier than the 0 dB TMR trials ($M = 70$ RAU, 68% correct, $SD = 24$ RAU), which were easier than the -6 dB TMR trials ($M = 58$ RAU, 58% correct, $SD = 22$ RAU). There were also main effects of Accent [$F(3, 120) = 17.26$, $p < 0.001$] and Gender [$F(1, 40) = 219$, $p < 0.001$], and these two factors also interacted significantly [$F(3, 120) = 7.7$, $p < 0.001$]. There was also a two-way interaction between Gender and TMR [$F(2, 80) = 15.12$, $p < 0.001$], and a three-way interaction between Accent, Gender, and TMR was also significant [$F(6, 240) = 4.44$, $p < 0.001$].

To better understand these interactions, separate analyses for each TMR trial type were conducted.

3.1 +6 dB TMR

The +6 dB TMR trials were subjected to a repeated-measures ANOVA with Accent and Gender as independent variables. There were main effects of Gender [$F(1, 40) = 37$, $p < 0.001$] and Accent [$F(3,120) = 9.2$, $p < 0.001$]. Listeners were more accurate on different gender trials (*M* = 101 RAU, 90% correct, *SD* = 22 RAU) than same gender trials (*M* = 89 RAU, 83% correct, *SD* = 22 RAU), indicating it is easier to selectively attend to voices that differ in terms of talker gender. Tukey tests established that listeners were more accurate on British-Canadian trials compared to British-British ($p < 0.02$) and Canadian-British ($p < 0.001$) trials (see Figure 3). This suggests that listeners were able to selectively attend to the British target talkers when they were paired with Canadian maskers, but they were less able to ignore the British talkers as maskers when they were paired with Canadian targets. Listeners were also more accurate on Canadian-Canadian trials than Canadian-British trials ($p < 0.02$), indicating, again, that listeners had a harder time ignoring British-accented maskers compared to Canadian-accented maskers.

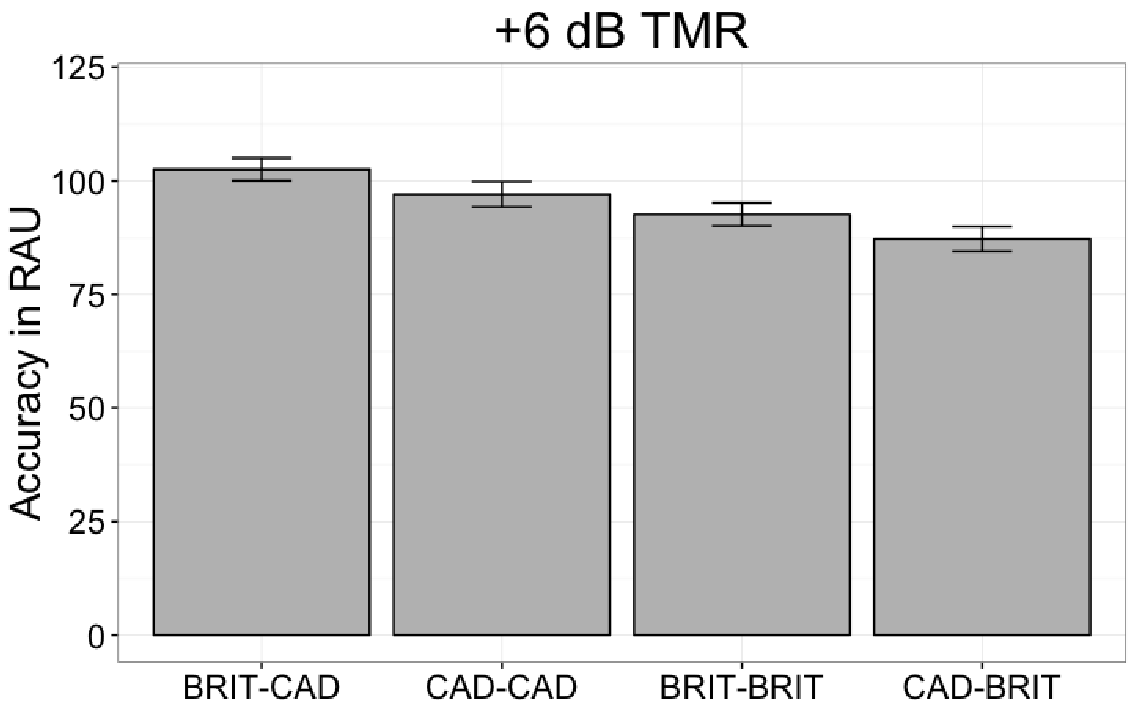

Figure 3. Proportion of coordinates from the target talker correctly identified on the +6 dB target-to-masker ratio trials as a function of target-masker Accent. BRIT = British-accented speaker, CAD = Canadian-accented speaker. Accent pairings are in the form of target-masker accent.

3.2 0 dB TMR

An identical analysis for the 0 dB TMR trials found an effect of Gender [$F(1, 40) = 208$, $p < 0.001$], Accent [$F(3, 120) = 8.2$, $p < 0.001$], and an interaction between Gender and Accent [$F(3, 120) = 2.8$, $p < 0.05$]. Listeners were less accurate on same gender trials ($M = 57$ RAU, 57% correct, $SD = 21$ RAU) than different gender trials ($M = 83$ RAU, 80% correct, $SD = 20$ RAU).

A Tukey test on Accent demonstrated that, like on the +6 dB TMR trials, listeners were more accurate on British-Canadian trials than Canadian-British ($p < 0.01$) and British-British ($p < 0.05$) trials at 0 dB TMR. This suggests that listeners were able to selectively attend to British voices as targets when they were paired with Canadian maskers, but were less effective at ignoring the British voices as maskers when they were paired with Canadian targets or British targets.

To understand the Gender by Accent interaction for the 0 dB TMR trials, separate analyses for the comparatively easy different gender trials and the harder same gender trials were run. A repeated-measures ANOVA for both different gender [$F(3, 120) = 7.3$, $p < 0.001$] and same gender trials [$F(3, 120) = 3.7$, $p < 0.05$] revealed an effect of Accent. These results are shown in Figure 4. For the different gender trials, Tukey tests revealed listeners were better at British-Canadian trials than British-British trials ($p < 0.01$) and Canadian-British trials ($p < 0.05$). Listeners were also more accurate on Canadian-Canadian trials than British-British trials ($p < 0.05$). For the more difficult same gender trials, a Tukey test showed that listeners were more accurate on British-Canadian trials than Canadian-British or Canadian-Canadian trials (both $p < 0.05$).

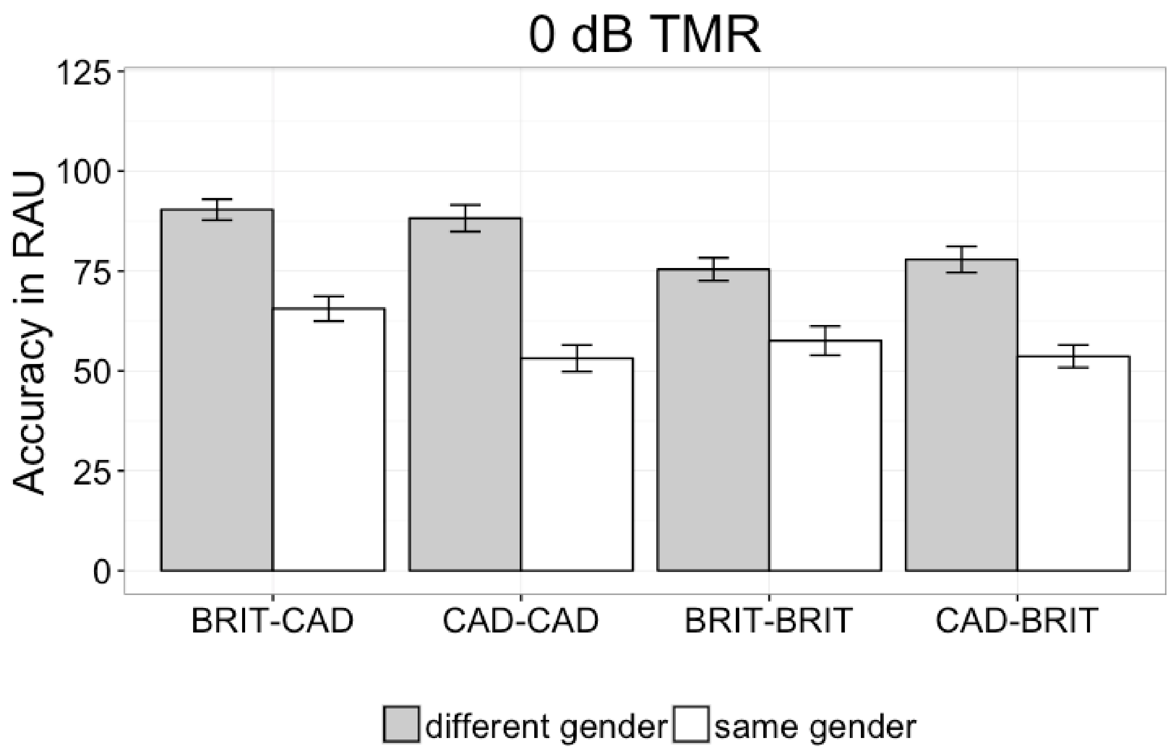

Figure 4. Proportion of coordinates from the target talker correctly identified on the 0 dB target-to-masker ratio trials as a function of target-masker Accent and Gender.

3.3 -6 dB TMR

The same analysis approach was taken for the -6 dB TMR trials. In this most challenging condition, there was a main effect of Gender [$F(1,40) = 92$, $p < 0.001$] and an interaction between Accent and Gender [$F(3, 120) = 15$, $p < 0.001$]. As in the +6 dB and 0 dB TMR conditions, listeners were more accurate on different gender trials ($M = 67$ RAU, 66% correct, $SD = 23$ RAU) than same gender trials ($M = 50$ RAU, 50% correct, $SD = 18$ RAU). The interaction between Gender and Accent is shown in Figure 5. To understand this interaction, separate analyses were run on different and same gender trials, with Accent surfacing as a main effect in both different [$F(3, 120) = 9.1$, $p < 0.001$] and same [$F(3, 120) = 7.2$, $p < 0.001$] gender

ANOVAs. Tukey tests confirmed that within the different gender trials, listeners were most accurate at talker segregation on Canadian-Canadian trials, significantly more so than the British-British and Canadian-British trials (both $p < 0.01$). On same gender trials, however, listeners were significantly less accurate on Canadian-Canadian trials compared to British-Canadian ($p < 0.01$) and Canadian-British ($p < 0.05$) trials. This indicates that during these most challenging trials (i.e., -6 dB TMR with no gender differences between the talkers), listeners perform better when the two talkers have different accents.

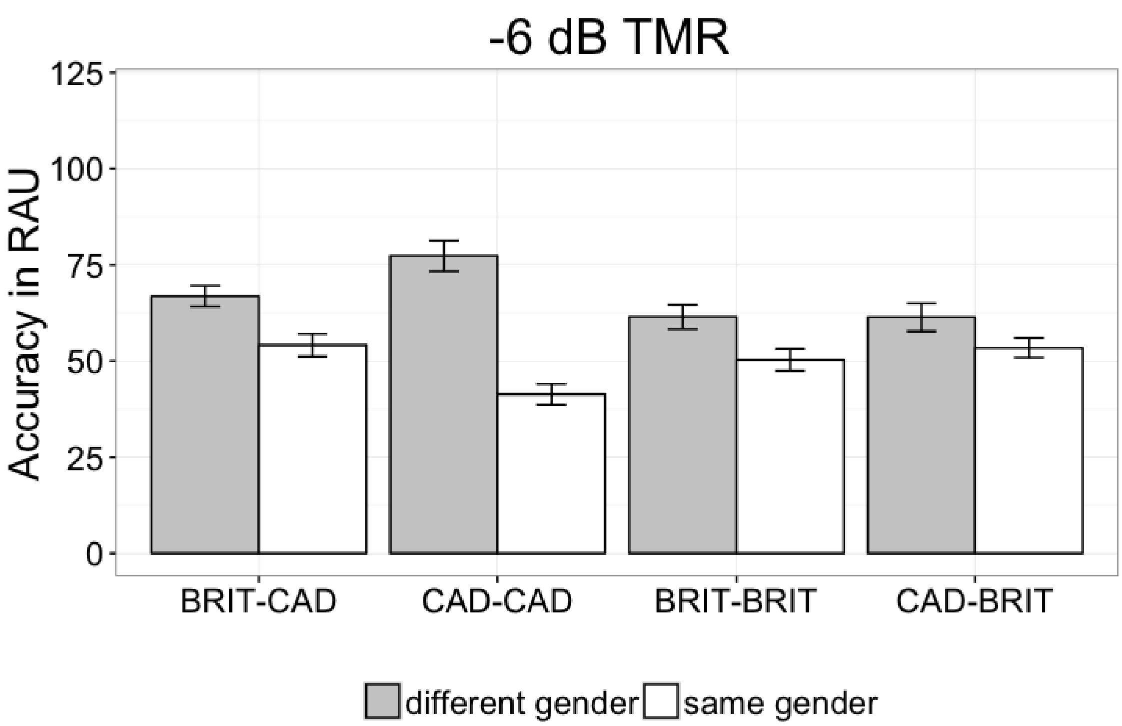

Figure 5. Proportion of coordinates from the target talker correctly identified on the -6 dB target-to-masker ratio trials as a function of target-masker Accent and Gender.

3.4 Effect size of Gender across the TMRs

In order to further examine the Gender effect, a series of paired t-tests were conducted comparing listeners' accuracy on different gender and same gender trials at each of the three TMRs. As established in the analysis above, there was a significant effect of Gender at each TMRs; performance on different gender trials was consistently higher than performance on same gender trials (see Table 2). The effect size of Gender was large at all TMRs.[4] However, the precise magnitude of the Gender effect varied across the three TMRs, with Gender having the largest influence at 0 dB TMR (d = 2.08), followed by -6 dB TMR (d = 1.12) and then +6 dB TMR (d =0.91). This suggests that gender cues are most effective when listeners cannot rely on intensity differences between the talkers as a cue to segregation.

Table 2

Mean accuracy on different gender and same gender trials by TMR

	Different gender trials		Same gender trials				
TMR	*M*	*SD*	*M*	*SD*	*t*	*df*	*d*
+6 dB	101.10	21.86	88.64	22.06	6.11***	40	0.91
0 dB	82.97	20.29	57.49	21.11	14.42***	40	2.08
-6 dB	66.74	22.55	49.82	18.48	9.62***	40	1.12

Note. Mean and standard deviations are in RAUs.
*** $p < 0.001$, one-tailed

4. Discussion

Performance in the speech segregation task varied greatly depending on the intensity difference between the target and masker talkers. This indicates that the three TMR levels represent three

[4] Cohen (1988) defines a large effect size as $d \geq 0.80$.

different listening conditions. Performance declined as intensity of the target talker decreased, with +6 dB TMR being easier than the 0 dB TMR and both +6 and 0 dB TMR being easier than -6 dB TMR. That is, regardless of the gender or accent of the two speakers, it was easiest for a listener to attend to the more acoustically intense of two signals and hardest for them to ignore a more intense signal in favour of a less intense one. Thus, while TMR can be a helpful cue to speech segregation, the effect of TMR is asymmetrical. A listener can use intensity differences to attend to the louder of two talkers but not to the quieter of the two. Unlike some results which indicate a release from masking at negative TMRs when the two talkers are identical (e.g., Brungart, 2001b; Brungart, 2005), performance always suffered at the negative TMR. This is likely because even when two speakers are the same gender or same accent, there are still inherent differences in the vocal characteristics of the two talkers that act as cues to segregation. It seems that only when the two talkers are identical – as in Brungart et al. (2001b, 2005)'s same talker trials – is a listener able to reliably attend to the quieter talker.

Our analyses also reveal that the cues a listener relies on may change depending on the listening environment. While the role of Gender remained relatively stable throughout the three TMRs, with different gender trials being consistently easier than same gender trials, the role of Accent varied. In a relatively easy listening situation (i.e., +6 dB TMR), listeners rely on differences in the two talkers' gender, with trials where the two speakers were a different gender (i.e., male-female or female-male trials) being easier than same gender (male-male or female-female) trials. Listeners can also use the talkers' accents as a cue to segregation on these trials. Regardless of gender, listeners perform best when asked to attend to a British-accented talker and ignore a Canadian-accented talker. Crucially, unlike the Gender effects, this is not due to the fact that the two talkers are speaking with different accents; rather, this effect has to do with the particular target-masker accent combination. Listeners do significantly worse when the target is a Canadian-accented speaker against a British-accented masker. Moreover, listeners find these Canadian-British trials harder than Canadian-Canadian trials, where there is no difference in target-masker accent. This suggests that an unfamiliar, British-accented talker is a more effective masker than a familiar, Canadian-accented talker. Thus, the unfamiliar British accent seems to have some salient characteristics that make it hard for listeners to ignore, even when it is the quieter of the two talkers.

In a moderately difficult listening situation (i.e., during 0 dB TMR trials), overall, listeners still do best when asked to attend to a British-accented speaker and ignore a Canadian-accented speaker but do not perform similarly well in the reverse scenario. Here, however, we start to see the role of Accent interact with Gender. Given two talkers with different genders, listeners are more likely to attend to a British talker than a Canadian talker, as was the case for the +6 dB TMR trials. They also continue to do worse at ignoring the unfamiliar accent than the familiar accent, as shown by the fact that they do better on British-Canadian trials than British-British or Canadian-British trials (i.e., trials with a British-accented masker). Here, we see an effect that was not present during the +6 dB TMR trials; whereas on the +6 dB trials listeners seemed to perform equivalently on same accent trials (that is, British-British and Canadian-Canadian trials), now listeners do better on the Canadian-Canadian trials than the British-British trials. Thus, it would appear that given two simultaneous talkers with the same intensity and accent, listeners are better able to exploit gender differences to separate the talkers when the two talkers are speaking with a familiar accent than when they are speaking with an unfamiliar accent.

On same gender, 0 dB TMR trials, we start to see fewer effects of Accent. Listeners do best on British-Canadian trials, but beyond this the target-masker accent combinations do not greatly impact performance. We know that the same gender trials are far more difficult than the different gender trials, and as the listening situation increases in difficulty (i.e., TMR decreases), Accent appears to be functioning as a less meaningful cue for segregation. This may be because having separate genders not only alleviates some of the difficulty of the task, but it may also reduce the cognitive load listeners face when trying to separate the two talkers. In turn, this might free up cognitive resources, allowing listeners to pay attention to and utilize accent cues. It does not seem to be the case that in the absence of gender cues, listeners default to using accent cues. Rather, it appears that on these more difficult 0 dB TMR, same gender trials, listeners struggle to use accent cues.

On -6 dB TMR trials, listeners continue to use Gender as a cue to segregation. On the easier different gender trials, listeners continue to struggle when the masker has a British accent. Now, this unfamiliar accent is made even more salient when in masker position, as on these trials it is the louder of the two talkers. Like during the 0 dB trials, listeners find it easier to exploit gender cues when two same accent talkers are Canadian (i.e., Canadian-Canadian trials) rather than British (i.e., British-British trials). Interestingly, on the hardest trials – same gender, -6 dB TMR trials, where not only does a listener have to ignore the more intense of two talkers but they no longer have any gender cues to aid them – listeners begin to use accent cues differently. Whereas on all other trial types the particular type of target-masker accent combination seemed to matter, now listeners *do* seem to rely on the speakers having different accents, with performance on *both* British-Canadian and Canadian-British trials surpassing that on Canadian-Canadian trials.

In this way, it seems evident that while listeners seem to use gender cues consistently across the three TMRs, listeners are not always making use of accent cues in the same way. On easy and moderately difficult trials, there seems to be a novelty effect. Listeners seem to benefit from the variation an unfamiliar accent offers; they easily pay attention to the unfamiliar accent and ignore the familiar accent, but struggle to ignore this unfamiliar accent. When the listening situation is challenging (-6 dB TMR) but the talkers are different genders, listeners benefit from having two familiarly-accented talkers and are better able to exploit gender cues in a familiar accent. On the most challenging trials (no gender cues at -6 dB TMR), listeners resort to distinguishing the two talkers in any way possible, and target-masker accent similarity rather than familiarity becomes most important. A British masker is no longer so detrimental to performance; instead, listeners seem to rely on having at least one British talker as a strategy to separate the two talkers, and they struggle to keep apart two familiarly-accented talkers. This finding is similar to the one found by Joshi et al. (2010), where listeners were able to use accent differences to separate multiple simultaneous talkers; when one talker had a different accent from the other talkers, listeners were better able to listen to that talker than to a talker who shared an accent of the other talkers (i.e., the “group” accent). Therefore, in challenging listening situations, listeners appear to rely on target-masker accent differences to distinguish speakers.

The results of this study also partially accord with the findings from Johnsrude et al. (2013), where listeners found it easier to selectively attend to a novel voice when the masker was familiar than when it was novel, except for on the less demanding trials with positive TMRs where no such benefit occurred. This is similar to the finding here that listeners were generally better at the British-Canadian trials than British-British trials. However, Johnsrude et al. (2013) also found that at all TMRs, listeners did best on familiar target trials, indicating that listeners could also easily selectively attend to their spouses' voice and ignore an unfamiliar talker of the same gender. This was not the case in the present study, where listeners generally struggled to tune out the unfamiliar British accent, except for on the most challenging trials. One potential reason for these discrepant findings is that familiarity with an individual's voice (i.e., one's spouse) may operate differently than familiarity with an entire accent. Perhaps the presence of a specific, highly familiar voice provided Johnsrude et al. (2013)'s listeners with such a boost in performance that it counteracted any negative effects associated with an unfamiliar masker.

It is important to note that the effect of Accent is separate from the effect of Gender. Presumably the Gender effect arose in part from the differences in fundamental frequency between male and female talkers, although there are also gender-based differences in vocal tract resonance and voice quality (e.g., Coleman, 1971) that likely also contributed. However, it is not the case that the effect of Accent is due to absolute differences in fundamental frequency between the two accents, as the fundamental frequencies of males and females across the two accents were comparable. This does not, however, preclude accent-specific differences in the overall pitch pattern of the phrases which listeners might have exploited. Also important is the fact that differences between the two accents were not due to differences in intelligibility. Only highly intelligible sentences – as determined in a speech-in-noise pretest – were used in the task. Thus, any signal-based differences between the two accents were due to differences in pronunciation.

It is interesting that the effect of Gender remained relatively consistent across the three TMRs whereas the effect of Accent did not; as the TMR decreased, the various target-masker accent combinations impacted listeners' performance less. The stability of the effect of target-masker gender differences may be due to the fact that listeners generally have lots of experience with listening to male and female talkers in everyday life. However, they may have more limited experience using accent differences as a means of distinguishing talkers. As follows, they may have more practise using Gender as a way to identify talkers, and gender cues may be more available to listeners than accent cues. This may be reflected in the fact that while Gender as mostly a signal-driven effect, Accent involves a combination of signal- and knowledge-driven factors; that is, in addition to accents being a property of the speech signals, their interpretation may largely depend on a listener's experience with them. It seems to be the case that while signal-driven effects like Gender are robust across different TMRs, listener-driven effects are less reliable and more susceptible to a listener's experience using the relevant cue.

An examination of the size of the Gender effect at each of the TMRs revealed that although Gender played an important role at each of the TMRs, it had the greatest influence on the 0 dB TMR trials. The effect size of Gender on the 0 dB TMR trials was approximately twice as large as on the other TMR trial types. In this way, there appears to be a tradeoff between Gender and TMR. When a listener can use intensity differences as a cue (i.e., on +6 and -6 dB TMR trials), Gender has a smaller effect than it does when there are no intensity differences (i.e., on 0 dB TMR trials). Thus, listeners seem to apply different listening strategies depending on which cues are available in the signal at any given moment.

Although I have discussed the results in relation to target-masker intensity and gender differences, my primary goal was to investigate the role of accents in a competing talker task. There has been some research on different accents in multi-talker situations (e.g., Joshi et al., 2010; Sinatra, 2012), but no clear consensus as to the role of accents in these situations. In this experiment, I directly compared a familiar accent (Canadian-accented English) to an unfamiliar accent (British-accented English). I found that listeners do differ in their ability to attend to these accents, and the ways in which listeners use unfamiliar and familiar accents as cues to talker segregation depend on the other acoustic cues (i.e., gender, intensity) that are available at any given moment. These results differ from those reported by Sinatra (2012) where listeners seemed to attend to a familiar accent (American-accented English) and to an unfamiliar accent (British-accented English) equally. It is important to note, however, the methodological differences between our studies that may underlie these differences in results. Sinatra (2012) used shadowing in the context of the dichotic listening task to investigate how much information from an unattended channel (i.e., the masker) was reported depending on the accent of the talker. In my experiment, I used the CRM, and my dependent measure was the proportion of trials on which the listener reported the correct coordinate from the *target*. Still, if an unfamiliar accent is harder to ignore than a familiar accent, as the results of my study appear to suggest, then it seems reasonable to expect that listeners in Sinatra (2012)'s study should have reported more information when the British-accented talker was in the unattended channel than when the masker spoke American-accented English. Regardless, the results of our studies converge on one important point: it is not always harder to pay attention to an unfamiliar accent than to a familiar accent when more than one individual is talking at the same time, and unfamiliar accents should not be regarded as necessarily “acoustically degraded.” Rather, the very variation that has been argued (e.g., Van Engen & Peelle, 2014) to render foreign-accented speech as “acoustically degraded” may prove beneficial under certain circumstances. That said, there are degrees of variation, and it likely is not the case that variation is categorically beneficial or detrimental to speech perception. The British accents in this study, while unfamiliar to the listeners, were still native English accents and were determined in the pretest to be highly intelligible. If stronger foreign accents were used – or if baseline levels of intelligibility were not controlled for – it is possible that the variation from an unfamiliar accent might be less helpful in a challenging listening environment.

It is also important to consider an alternative explanation for these results. Although I have discussed the primary difference between the accents in this study as being one of familiarity and this explanation seems likely given the findings of Johnsrude et al. (2013), it is also possible that social prestige may be involved. Sumner (2015) argues that a listener’s social biases can influence how much attention they give a talker and therefore how likely the listener is to encode and retain information from that talker; in general, listeners remember more from esteemed accents than those that are less prestigious. If the Canadian listeners in my study viewed the British accents as more socially prestigious than the Canadian accents, then the increased desirability of the British accents may be at least partially responsible for the listeners’ increased attention to the British-accented talkers on the easier trials. Future research might seek to tease apart the effects of familiarity and social desirability by incorporating an unfamiliar accent that is known to be less socially prestigious than the local accent but still equally intelligible.

My study has some key limitations. Although the CRM task permits a high degree of control over aspects such as sentence content and syntax, its scripted nature makes it fundamentally unnatural. This may limit the ability to apply these results to complex, real-world situations where more than one individual is talking. Future studies may consider implementing methods such as Hafter et al. (2013)'s naturalistic cocktail party method to better approximate the demands of real-world listening situations. Additionally, our understanding of accents as a speech segregation cue would benefit from investigations of multiple simultaneous talkers each with a different accent.

5. Conclusion

The results of this study contribute to our understanding of the factors that enable one to understand one person's speech against a background of other voices (i.e., the cocktail party problem). While the prediction that listeners would find it more challenging to pay attention to an unfamiliar accent than a familiar one was not supported, these results do demonstrate that listeners may use target-masker accent as a cue to speech segregation. I presented two hypotheses of how listeners might approach unfamiliar accents in a competing talker situation. On the whole, these results support the latter hypothesis; that is, the variation offered by the unfamiliar accents often had a facilitative effect on performance. Specifically, listeners appeared to easily direct their attention to an unfamiliar British accent on all but the most difficult trials. Whether this effect is specific to the accents used in the study or can be applied to British accents or unfamiliar accents in general is unknown at the current time, and future research should seek to elucidate the mechanism behind this finding.

Acknowledgements

This project would not exist without the help of many individuals. A huge thank you to my supervisor, Dr. Molly Babel, for encouraging me to pursue an honours thesis and for her guidance, inspiration, and encouragement every step of the way! I would also like to thank Martin Oberg for sharing his Python expertise and for help with stimuli preparation, the Canadian-accented speakers who so enthusiastically volunteered their voices to this project, and all the wonderful members in the Speech in Context Lab for their ongoing support.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 520-529.

Boersma, P. & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.21) [computer software]. Available from http://www.praat.org/

Bolia, R. S., Nelson, W. T., Ericson, M. A., & Simpson, B. D. (2000). A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America*, *107*(2), 1065-1066.

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica, 86*(1), 117-128.

Brungart, D. S. (2001a). Evaluation of speech intelligibility with the coordinate response measure. *The Journal of the Acoustical Society of America*, *109*(5), 2276-2279.

Brungart, D. S. (2001b). Informational and energetic masking effects in the perception of two simultaneous talkers. *The Journal of the Acoustical Society of America*, *109*(3), 1101-1109.

Brungart, D. S. (2005). Informational and energetic masking effects in multitalker speech perception. In P. Divenyi (Ed.), *Speech separation in humans and machines* (pp. 261–267). New York: Springer

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., & Kidd Jr, G. (2005). Across-ear interference from parametrically degraded synthetic speech signals in a dichotic cocktail-party listening task. *The Journal of the Acoustical Society of America*, *117*(1), 292-304.

Brungart, D. S., Simpson, B. D., Ericson, M. A., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*,*110*(5), 2527-2538.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*(5), 975-979.

Clarke, S., Elms, F., & Youssef, A. (1995). The third dialect of English: Some Canadian evidence. *Language Variation and Change*, *7*(2), 209-228.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Rutledge.

Coleman, R. O. (1971). Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech, Language, and Hearing Research*, *14*(3), 565-577.

Cooke, M., Barker, J., Cunningham, S., & Shao, X. (2006). An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America*, *120*(5), 2421-2424.

Cooke, M., Garcia Lecumberri, M. L., & Barker, J. (2008). The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception. *The Journal of the Acoustical Society of America*, *123*(1), 414-427.

Ericson, M. A., Brungart, D. S., & Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *The International Journal of Aviation Psychology*, *14*(3), 313-334.

Ericson, M.A., & McKinley, R. L. (1997). The intelligibility of multiple talkers separated spatially in noise. In R. H. Gilkey & T. R. Anderson (Eds.), *Binaural and Spatial Hearing in Real and Virtual Environments* (pp.701-724). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Garcia Lecumberri, M. L., & Cooke, M. (2006). Effect of masker type on native and non-native consonant perception in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445-2454.

Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., Cohen, J. I., & Waldroup, C. (2013). Recognition of accented and unaccented speech in different maskers by younger and older listeners. *The Journal of the Acoustical Society of America*, *134*(1), 618-627.

Hafter, E. R., Xia, J., & Kalluri, S. (2013). A naturalistic approach to the cocktail party problem. In Moore, B. C. J., Patterson, R. D., Winter, I. M., Carlyon, R. P., & Gockel, H. E., *Basic Aspects of Hearing* (pp. 527-534). Springer: New York.

Johnsrude, I. S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H. P., & Carlyon, R. P. (2013). Swinging at a cocktail party voice familiarity aids speech perception in the presence of a competing voice. *Psychological Science*, *24*, 1994-2005.

Joshi, M., Iyer, M., & Gupta, N. (2010). Effect of accent on speech intelligibility in multiple speaker environment with sound spatialization. In *Information Technology: New Generations (ITNG), 2010 Seventh International Conference on* (pp. 338-342). IEEE.

Moore, T. (1981). Voice communication jamming research. In *Advisory Group for Aerospace*

Research and Development Conference Proceedings (No. 311, pp. 2:1-2:6).

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*(3), 289-306.

Peng, J., Zhang, H., & Wang, Z. (2012). Chinese speech identification in multi-talker babble with diotic and dichotic listening. *Chinese Science Bulletin*, *57*(20), 2548-2553.

Psychology Software Tools, Inc. (2012). E-Prime 2.0 [computer software]. Available from http://www.pstnet.com

REAPER: Robust Epoch And Pitch EstimatoR [computer software]. Available from https://github.com/google/REAPER

Rogers, C. L., Dalby, J., & Nishi, K. (2004). Effects of noise and proficiency on intelligibility of Chinese-accented English. *Language and Speech*, *47*(2), 139-154.

Sinatra, A. M. (2012). *The impact of degraded speech and stimulus familiarity in a dichotic listening task* (Doctoral dissertation, University of Central Florida Orlando, Florida).

Sinatra, A. M., Sims, V. K., Najle, M. B., & Bailey, S. K. (2012). The impact of synthetic and accented speech on unattended recall in a dichotic listening task. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1635-1638). SAGE Publications.

Sumner, M. (2015). The social weight of spoken words. *Trends in Cognitive Sciences, 19*(5), 238-239.

University of British Columbia Planning and Institutional Research Office (UBC PAIR). (2015). *2015 annual report on enrolment: Total enrolment summary.* Retrieved November 15, 2016 from http://pair2016.sites.olt.ubc.ca/files/2016/04/2015-Enrolment-Report.pdf

Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*, 577.

Wells, J. C. (1982). *Accents of English* (Vol. 2-3). Cambridge University Press.

Yan, Q., & Vaseghi, S. (2002, May). A comparative analysis of UK and US English accents in recognition and synthesis. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on* (Vol. 1, pp. I-413). IEEE.

Zeller, C. (1993). Linguistic symmetries, asymmetries, and border effects with a Canadian/American sample. In S. Clarke (Ed.), *Focus on Canada* (pp.179-199). Amsterdam: John Benjamins Publishing.