NeuroImage 181 (2018) 582–597

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural architecture underlying person perception from in-group and out-group voices

Xiaoming Jiang [a,*], Ryan Sanford [b], Marc D. Pell [a,b,**]

[a] School of Communication Sciences and Disorders, McGill University, Montréal, Canada
[b] McConnell Brain Imaging Center, Montréal Neurological Institute, McGill University, Montréal, Canada

ARTICLE INFO

Keywords:
Social perception
Vocal expression
Trust
In-group bias
fMRI
Functional connectivity

ABSTRACT

In spoken language, verbal cues (*what* we say) and vocal cues (*how* we say it) contribute to person perception, the process for interpreting information and making inferences about other people. When someone has an accent, forming impressions from the speaker's voice may be influenced by social categorization processes (i.e., activating stereotypical traits of members of a perceived 'out-group') and by processes which differentiate the speaker based on their individual attributes (e.g., registering the vocal confidence level of the speaker in order to make a trust decision). The neural systems for using vocal cues that refer to the speaker's identity *and* to qualities of their vocal expression to generate inferences about others are not known. Here, we used functional magnetic resonance imaging (fMRI) to investigate how speaker categorization influences brain activity as Canadian-English listeners judged whether they *believe* statements produced by in-group (native) and out-group (regional, foreign) speakers. Each statement was expressed in a confident, doubtful, and neutral tone of voice. In-group speakers were perceived as more believable than speakers with out-group accents overall, confirming social categorization of speakers based on their accent. Superior parietal and middle temporal regions were uniquely activated when listening to out-group compared to in-group speakers suggesting that they may be involved in extracting the attributes of speaker believability from the lower-level acoustic variations. Basal ganglia, left cuneus and right fusiform gyrus were activated by *confident* expressions produced by out-group speakers. These regions appear to participate in abstracting more ambiguous believability attributes from accented speakers (where a conflict arises between the tendency to disbelieve an out-group speaker and the tendency to believe a confident voice). For out-group speakers, stronger impressions of believability selectively modulated activity in the bilateral superior and middle temporal regions. Moreover, the right superior temporal gyrus, a region that was associated with perceived speaker confidence, was found to be functionally connected to the left lingual gyrus and right middle temporal gyrus when out-group speakers were judged as more believable. These findings suggest that identity-related voice characteristics and associated biases may influence underlying neural activities for making social attributions about out-group speakers, affecting decisions about believability and trust. Specifically, inferences about out-group speakers seem to be mediated to a greater extent by stimulus-related features (i.e., vocal confidence cues) than for in-group speakers. Our approach highlights how the voice can be studied to advance models of person perception.

1. Introduction

Our perception of other people is often shaped by *how* they speak, i.e., by information derived from the speaker's voice. Vocal cues in speech provide simultaneous cues about the speaker's identity (e.g., sex, social class, linguistic/geographical background, Belin et al., 2004; Campanella and Belin, 2007) and about the speaker's affective and mental state (e.g., that the speaker is confident or doubtful in what they are saying, Jiang and Pell, 2015, 2017). Research shows that these different sources of vocal information are rapidly processed to elaborate their social relevance to the listener (e.g. Bresnahan et al., 2002; Kotz et al., 2006; Schirmer, 2017). However, little is known about the *interplay* of vocal cues that refer to a speaker's identity and that express interpersonal meanings in spoken language, nor the neural substrates that coordinate

* Corresponding author. Department of Psychology, Tongji University, 1239 Siping Road, Shanghai, China.
** Corresponding author. School of Communication Sciences and Disorders, McGill University, 2001 McGill College, 8th Floor, Montréal, Canada.
E-mail addresses: xiaoming.jiang@tongji.edu.cn (X. Jiang), marc.pell@mcgill.ca (M.D. Pell).

https://doi.org/10.1016/j.neuroimage.2018.07.042
Received 5 September 2017; Received in revised form 4 July 2018; Accepted 16 July 2018
Available online 19 July 2018

and integrate these different cues during person perception.

The current study addressed this gap in the literature by looking at how listeners process vocally-expressed confidence in the context of 'out-group' voices– defined here as speakers who possess an accent that is perceived as different by the listener. Specifically, we investigated how confidence- and accent-related information jointly influences decisions about whether to *believe* what a speaker says. These data will build constructively on recent descriptions of the neural substrates used to generate trust/believability decisions when listeners hear 'in-group' speakers with the same accent (Jiang et al., 2017).

1.1. Speaker categorization and accent: behavioral and neural evidence

Voice-based person perception is likely to differ when social partners have an accent–a manner of pronunciation that affects the phonetic and prosodic rules employed by a particular speech community (Lippi-Green, 1997). Accents quickly mark whether a speaker shares group membership with the listener (in-group voice), and thus, whether the speaker is mentally assigned out-group status due to perceived differences in their first language and/or geographical background (Bresnahan et al., 2002). Behavioral studies have demonstrated ways that accented speech alters linguistic comprehension (Adank et al., 2009; Dupoux and Green, 1997; Derwing and Munro, 1997) and influences social decision-making processes (Bresnahan et al., 2002; Hayakawa et al., 2016; Hu et al., 2012; Sumner, 2015). Recent fMRI data comparing accents in English suggest that there is an *affective bias* towards the native accent (the preference of one's own accent to other varieties of one's native language), yielding distinct patterns of cerebral activation for in-group versus out-group speakers (Bestelmeyer et al., 2015). Using an adaptation paradigm, the authors reported that brain activity in bilateral amygdalae, right rolandic operculum (extending to right superior temporal gyrus (STG)), and anterior cingulate cortex (ACC) increased when participants listened to an accent similar to their own (south England accent), whereas it decreased when exposed to a less familiar accent (Scottish or American speaker). These observations imply that greater social salience is assigned, or more attention is deployed, to one's own accent, contributing to differences in how in-group and out-group voices are processed by the brain (Bestelmeyer et al., 2015). However, stimuli presented in the previous study were lists of digits produced by speakers in a neutral tone of voice. Could having an out-group accent alter the salience of other vocal cues that encode interpersonal meanings in speech, such as whether or not to *believe* what a speaker is saying based on their voice?

Major theories of social categorization (e.g. Bresnahan et al., 2002; Fiske and Neuberg, 1990) predict that speaker accent information could affect neural responses underlying person perception in a variety of ways. Group membership causes favoritism towards in-group over out-group individuals, promoting a 'native accent bias' in the social evaluation of speakers. For example, listeners typically judge their own accent, or an accent similar to their own, as more favorable (Ryan and Sebastian, 1980) and trustworthy (Lev-Ari and Keysar, 2010). Listeners can also categorize social characteristics of a speaker, such as race, more accurately when listening to their own accent (Perrachione et al., 2010; Kuhl et al., 1992). Differences in the social categorization of accented speakers have been linked to processes for extracting dialectal features of the voice (Perrachione et al., 2010), which activate socially-acquired knowledge associated with members of the perceived out-group, affecting social perception. This in-group bias may reduce empathy-related neural responses towards individuals judged to be a member of another group, as shown when in-group/out-group participants were subjected to a negative stimulus (increased activation in ACC and inferior frontal gyrus (IFG) were reported when viewing an in-group face receiving a painful stimulus, Xu et al., 2009; Sheng and Han, 2012). This suggests that stereotypical information is activated during social categorization and this knowledge affects concurrent neurocognitive processes more generally (e.g. monitoring conflicting social information, Hehman et al., 2014).

1.2. Effects of group status on vocal confidence and believability

Given evidence of a native-accent bias in social perception, categorizing speakers as belonging to an out-group is likely to modulate neural regions for generating broader social inferences from a speaker's voice, such as whether or not to believe what a speaker is saying. Vocally-expressed confidence— dynamic changes in the pitch, loudness, and rate of speech that encode meaning along a gradient from "confident" to "doubtful"— carries salient information regarding whether a speaker is believable or can be trusted (Jiang and Pell, 2015, 2016a). When listeners perceive a speaker to be more confident, their voice shows graded increases in mean intensity and intensity variation, and their statements tend to be produced more quickly and with a low/falling voice pitch, among other possible acoustic parameters (Jiang and Pell, 2017).

The neural correlates underlying vocal confidence processing, and how this information is used when making believability decisions about in-group speakers, were recently examined in an fMRI design (Jiang et al., 2017). English-Canadian participants listened to short statements varying in vocally-expressed confidence produced by native speakers and were asked to form believability impressions about each statement. Neural activities associated with different vocal expressions (confident, doubtful, neutral) and with particular believability responses (highly- or lowly-believed statements) were then analyzed. Perceptually, statements such as *She'll do a good job,* produced in a confident or neutral tone, were rated as more believable than the same statement expressed in a doubtful manner. At the neural processing level, confident voices tended to be associated with more brain activations in the left superior frontal gyrus (SFG), left IFG, and right supplementary motor area (SMA) compared to doubtful voices. Meanwhile, doubtful voices were associated with increased activation in the right superior temporal gyrus when compared to confident and neutral expressions. These data suggest that as listeners make believability decisions, they may attend to vocal confidence information provided by the speaker, and that the type of vocal expression most likely differentially activates regions in a fronto-temporal network dedicated to prosody (voice) perception (e.g., Hellbernd and Sammler, 2018; Sammler et al., 2015).

Interestingly, evaluating believability about in-group statements produced in a neutral tone, which lacked explicit or intentional vocal markers of confidence or doubt, was associated with stronger activity compared to other vocal expressions in the bilateral cerebellum, bilateral cuneus, and right fusiform. These regions are commonly viewed to be part of mentalizing and mirroring networks, which are functional networks that are hypothesized to play a key role in social interactions (Van Overwalle, 2009; Van Overwalle et al., 2014; Van Overwalle and Baetens, 2009) and are observed in situations when social attributions are demanded (e.g. Kuhlen et al., 2015). Notably, greater involvement was observed in: anterior inferior parietal sulcus (aIPS), a region which provides rapid and intuitive readings of a speaker's nonverbal behavior and helps recognize the goal of a perceived action (or the "mirroring" process); and cerebellum, a region that may help to abstract and mentalize another person's behavior when vocal cues are ambiguous (or the "mentalizing" process; Baetens et al., 2013; Van Overwalle et al., 2014). In terms of the actual believability responses, whenever in-group speakers were judged to be more believable, we observed greater neural activity in right superior parietal lobule (SPL) and post-central gyrus (PoCG), whereas activation increased in the left PoCG/SPL as speakers were judged to be less believable. The right PoCG/SPL was interpreted as mediating the inference that speakers are more believable from vocal cues such as pitch height and spectrotemporal regularity (Bestelmeyer et al., 2012). The left PoCG/SPL was associated with recognition of speech acts with a disposition of lacking trustworthiness on the part of the speaker (Jiang et al., 2017). The activity from the right SPL synchronized with the ACC and right medial SFG, and the activity from the left PoCG synchronized with the ACC and dorsal medial prefrontal cortex(mPFC), suggesting heightened engagement of attentional control networks for speakers whose vocal expressions reveal salient cues regarding one's believability (Jiang et al., 2017). These results shed new light on neural systems for evaluating vocally-expressed confidence to form believability impressions about *native* speakers who belong to the same speech community as the listener.

Using vocal confidence information to form believability impressions is hypothesized to differ in the context of accents that mark the speaker's out-group status, due to the activation of out-group-related knowledge that biases how vocal attributes of the speaker are assigned relevance or 'individuated' (Fiske and Neuberg, 1990; Lev-Ari, 2014; Perrachione and Wong, 2007; Jiang et al., In Review). This hypothesis has received support from recent neurophysiological data highlighting the temporal interplay of speaker accent information and vocally-expressed confidence as listeners attend to in-group versus out-group voices (Jiang et al., In Review). In an event-related potential study, when listeners were presented with vocal confidence expressions produced by in-group and out-group speakers, the brain response at 200 ms towards the in-group speakers was more positive than the out-group speakers. Moreover, these positive responses significantly varied in confident expressions for out-group speakers with a regional or foreign accent. These results suggest that when rendering believability judgments, vocal expressions of confidence are given different contextual relevance by the neurocognitive system after categorizing the speaker as an out-group member (Jiang et al., In Review). These data provide initial evidence that the out-group voice moderates neural mechanisms underlying believability inferences by possibly altering the perceptual weight or significance accorded to particular types of vocal confidence cues. However, establishing how social categorization processes affect the *neural circuitry* involved when evaluating out-group versus in-group speakers (Jiang et al., 2017) has not yet been tested.

1.3. The present study

The objective of our study is to identify the neural mechanisms underlying the encoding of speaker confidence and inferring believability from individual statements, to determine how these mechanisms are moderated when the speaker and listener share the same accent or not. Our stimuli were short English statements expressed in a confident, doubtful, or neutral voice; the same statements were produced by native speakers of Canadian English ('in-group' voice) and speakers from two out-groups: those who speak a regional variant of English (Australian); and those who speak English as a second language (native speakers of French). "Out-group" speakers were defined based on the linguistic background and accent of the speaker and included those with both regional and foreign accents to permit more generalizable conclusions about the effects of out-group voices on interpersonal believability and trust. While our primary focus in this study was to compare effects due to in-group versus out-group voices, defining two types of out-groups also allowed certain factors, such as accent intelligibility, attitudes towards particular accents, etc., to be usefully compared in our analyses.

We are interested in the following questions. First, does the out-group status of a speaker affect believability judgments about statements, and if so, are in-group voices judged to be more believable than speakers with out-group accents? Second, is the process of decoding a speaker's vocally-expressed confidence level associated with qualitatively distinct neural activities when the speaker carries an "out-group" vs. "in-group" accent? To resolve this question, we built contrasts between vocal expressions and interaction models of vocal expression and speaker accent. We hypothesized that speaker accent would have a significant impact on the neural activities that are involved as listeners make believability inferences from vocal expressions; for example, out-group accents may differentially bias how attention is deployed to vocal confidence expressions, affecting their perceived motivational significance to the listener (Jiang et al., In review). Whereas increased activity was observed for neutral statements produced in a neutral tone by in-group speakers, implicating the mentalizing and mirroring networks (Jiang et al., 2017), we might expect out-group voices to increase related neural activity (e.g. bilateral cuneus, fusiform, cerebellum, etc.) for *confident* expressions (Jiang et al., In Review). This condition may place greater demands on processes for mentalizing the speaker because explicit confidence cues contradict an initial negative expectation about speaker believability activated when categorizing the out-group voice (Van Overwalle, 2009; Van Overwalle and Baetens, 2009).

Third, we sought to discover whether inferring believability about out-group speakers is linked to changes in regional activity and functional connectivity in brain regions that subserve the decoding of vocally-expressed confidence. To understand the connectivity, parametric models were built to regress individual believability ratings on the BOLD signals. We have shown previously that bilateral SPL/PoCG is involved in the believability evaluation, at least for in-group speakers (Jiang et al., 2017). Given that a speaker with an accent may increase attention to contextual cues, it is possible that bilateral temporal activations are involved to a relatively greater extent when evaluating out-group speakers (Gibson et al., 2017; Lev-Ari, 2014). When an accented speaker is categorized as an out-group member, listeners may be forced to rely more heavily on stimulus attributes furnished by the speaker during the mentalizing process, i.e., to conduct a more elaborate "piece-meal analysis" of the speaker's vocal confidence cues (Fiske and Neuberg, 1990; Perrachione et al., 2010). However, this process should be more effortful in the face of out-group accents, leading to increased medial temporo-occipital activities for extracting vocally-expressed confidence in out-group vs. in-group speakers (Hehman et al., 2014; Li et al., 2016). Connectivity patterns associated with the believability inference for different accent types (Stanley et al., 2012) were also explored for the first time, allowing new insights that elaborate models of vocal expression processing (e.g., Frühholz and Grandjean, 2013; Kotz et al., 2006) and mentalizing/inferring speech acts (e.g., Egorova et al., 2017; Kuhlen et al., 2015; Sammler et al., 2015; Van Overwalle and Baetens, 2009).

Finally, we posed the question: can the different neural activities involved in the evaluation of out-group accents be predicted by differences in a listener's favorability towards an accent or its perceived intelligibility? The perception of a speaker with the same accent promotes judgments of enhanced in-group favorability (Coupland and Bishop, 2007; Edwards, 1982; Hurt and Weaver, 1972; Mulac et al., 1974; Lev-Ari and Keysar, 2010; Ryan and Sebastian, 1980). Therefore, we hypothesized that accent-derived attitudes in individual listeners would modulate neural activities underlying the perception of conflicting social information (ACC, mPFC, e.g. to arrive at a positive judgment of an out-group speaker) or neural responses which index the greater social relevance of in-group vocal cues (e.g. amygdalae, Bestelmeyer et al., 2015).

2. Method

2.1. Participants

Twenty-six right-handed adults participated in the study (14 females/12 males; Age: 22.6 yrs, from 18-30 yrs; Years of Education: 15.9 yrs, from 13-19 yrs). All participants spoke Canadian English as their mother tongue, reported English as the language they used most in daily communications, and none had lived outside of Canada. Fifteen participants reported learning a second language (French, German, Cantonese) after five years of age. The study was conducted in accordance with the Helsinki Declaration, approved by the Research Ethics Board, Montreal Neurological Institute and Hospital, as well as by the Institutional Review Board, Faculty of Medicine, McGill University. Written informed consent was obtained from all participants who were compensated for their involvement.

2.2. Materials and design

English statements (n = 270), expressed in a confident, doubtful or neutral voice, were selected from an inventory of vocal confidence recordings (Jiang and Pell, 2017; Jiang et al., In review). All statements referred to personal knowledge held by the speaker (e.g., *She has access to the building*), meaning that only vocal cues could be used by listeners to infer speaker believability. Thirty identical statements were expressed in each tone of voice by Canadian-English speakers (1 Female/1 Male), speakers of a regional variant of English (Australian, 1 Female/1 Male), and speakers of a foreign language (Québécois-French, 1 Female/1 Male). Australian speakers were raised by English-speaking parents in (East Coast) Australia and moved to Canada less than one year prior to the recordings. Québécois speakers were raised by French-speaking parents in the Lac St-Jean region of Québec, Canada and had learned English after 10 years of age in a formal classroom setting; they were selected for having a high rate of fluency in English in the presence of a detectable accent to Canadian-English listeners (Jiang et al., in review). In the context of our study, Australian and French speakers are both considered to have 'out-group' accents or voices in relation to our listeners[1]; for certain analyses, we may also refer to the three accent conditions as exemplifying *Native, Regional*, or *Foreign* accents in English. All speakers with out-group accents used in the study were rated as having a manner of articulation that was significantly more distant from the Native accent by a group of Canadian listeners (Jiang et al., in review).[2] A total of 270 stimuli were used (30 items × 3 vocal expressions x 3 accents).

Table 1 provides descriptive details of the acoustic features of selected stimuli by vocal expression and accent type. We extracted the following measures: mean fundamental frequency (f0), f0 range, mean amplitude, amplitude range, utterance duration, and mean harmonics-to-noise ratio (HNR) (Jiang and Pell, 2017). In addition, while our experiment focuses on judgments of speaker believability from vocal cues, we gathered basic information on how listeners rated vocally-expressed confidence across accent types for the experimental stimuli. An independent group of 18 native Canadian-English listeners (Age: M = 21.4 yrs, range = 18–26; Education: M = 14.7 yrs, range = 13–18) who did not participate in the fMRI scanning rated how confident the speaker sounded on a 5-pt scale (1 = not at all, 5 = very much). The linear mixed effects model was used to verify whether the three types of vocal expressions were perceptually differentiated in a similar manner by speakers in the three accent groups, with the level of confidence as the fixed factor. The main effect of confidence was significant in all accents (Native: $F(2, 57) = 354.38$, $p < .0001$; Regional: $F(2, 57) = 228.35$, $p < .0001$; Foreign: $F(2, 57) = 124.55$, $p < .0001$). Mean ratings were higher for confident than doubtful expressions produced by the Native ($b = 2.59$, $t = 25.24$, $p < .0001$), Regional ($b = 1.88$, $t = 20.54$, $p < .0001$), and Foreign ($b = 1.49$, $t = 15.07$, $p < .0001$) accented speakers. Neutral expressions were perceived as relatively closer to confident than doubtful expressions in all three accents (see Table 1).

2.3. fMRI scanning parameters

The fMRI scanning and other relevant tests were performed at the Montréal Neurological Institute. Scanning was performed on the same 3-T Siemens Imager with a 32-channel head coil. Structural T1-weighted anatomical images were first acquired for anatomical reference for each participant (repetition time (TR)/echo time (TE)/inversion time (TI) = 2300/2.98/900 ms, flip angle = 9° and voxel size = 1 × 1x1mm^3). Six functional echo-planar runs were then acquired for each participant. Each functional run contained 41 slices with 53 vol with whole-head interleaved acquisition (TR/TE = 8000/30 ms, acquisition time per run = 7 min 14 s, flip angle = 90°, field of view (FOV) = 224 mm, image matrix = 64 × 64, distance factor = 0%, and voxel size = 3.5 × 3.5 × 3.5 mm^3). Volumes were acquired every 8 s and lasted for 2.5 s after the presentation of the vocal expression. Each functional scan used a sparse-sampling paradigm, which minimizes the influence of the BOLD response due to scanner noise (Belin et al., 1999). This paradigm takes advantage of the 4- to 6-sec delay in the hemodynamic response peak following the stimulus (Gaab et al., 2007).

Table 1
Major perceptual and acoustic parameters of the experimental stimuli (mean, standard deviation).

Measure	Vocal Expression		
	Confident	Doubtful	Neutral
	In-group accent		
Perceived Confidence (1–5)[a]	4.33 (.40)	1.74 (.51)	3.79 (.32)
Fundamental frequency Mean (Hz)[b]	.23 (.15)	.43 (.16)	.27 (.15)
Fundamental frequency Range (Hz)[b]	1.91 (1.61)	1.40 (.89)	1.36 (1.43)
Amplitude Mean (dB)[b]	.33 (.04)	.34 (.05)	.35 (.05)
Amplitude Range (dB)[b]	.66 (.09)	.64 (.15)	.57 (.11)
Harmonics-to-noise Mean (dB)	-.23 (.26)	.04 (.31)	.00 (.27)
Duration Mean (seconds)	1.65 (.23)	1.95 (.56)	1.40 (.27)
	Out-group/Regional accent		
Perceived Confidence (1–5)	3.83 (.36)	1.95 (.28)	3.36 (.45)
Fundamental frequency Mean (Hz)	.24 (24)	.39 (.17)	.21 (.14)
Fundamental frequency Range (Hz)	1.48 (1.52)	1.59 (.99)	1.39 (1.54)
Amplitude Mean (dB)	.40 (.03)	.40 (.03)	.40 (.03)
Amplitude Range (dB)	.61 (.09)	.59 (.09)	.62 (.12)
Harmonics-to-noise Mean (dB)	-.16 (.45)	.06 (.35)	-.01 (.50)
Duration Mean (seconds)	1.32 (.20)	1.56 (.32)	1.41 (.19)
	Out-group/Foreign accent		
Perceived Confidence (1–5)	3.48 (.52)	1.99 (.39)	3.14 (.44)
Fundamental frequency Mean (Hz)	.24 (.10)	.37 (.13)	.21 (.06)
Fundamental frequency Range (Hz)	1.13 (1.07)	1.60 (1.21)	1.20 (1.23)
Amplitude Mean (dB)	.41 (.03)	.41 (.03)	.41 (.03)
Amplitude Range (dB)	.67 (.06)	.66 (.06)	.66 (.06)
Harmonics-to-noise Mean (dB)	-.16 (.22)	.04 (.22)	.00 (.21)
Duration Mean (seconds)	1.54 (.29)	1.62 (.34)	1.62 (.26)

[a] Speaker confidence ratings were gathered from an independent group of 18 Canadian English listeners who did not participate in the fMRI experiment.
[b] To allow meaningful comparisons across speakers and recording sessions, fundamental frequency and amplitude values were first normalized for each statement and speaker to express each measure as its proportional distance from the speaker's own resting frequency/amplitude (i.e., the mean of each speaker's minimum frequency/amplitude in neutral statement, Jiang and Pell, 2017).

2.4. fMRI task procedure

Fig. 1 demonstrated the experimental procedure and the event flow of one trial in the fMRI task. Each vocal stimulus was presented during the silent periods between acquisitions. Stimuli representing different vocal expressions and accent conditions were fully randomized within runs. Listeners judged "*Do you believe the speaker*" along on a 4-pt scale (1 = not at all, 4 = very much). The vocal stimulus was preceded by a fixation of 0.5 s and followed by a 1 s rating scale signaling the listener to respond. There were no significant differences between accents in the duration of vocal statements (Native: 1.67 ±0 .45 s; Regional: 1.43 ±0 .26 s; Foreign: 1.59 ±0 .30 s, ps > .1). The onset of vocal stimuli was jittered so that the center point of each vocal expression was 4.25 s from the middle of the subsequent scanning period (Rodd et al., 2005). Thirty null events were randomly mixed with the vocal events (Bach et al., 2008). Null events

[1] Here, both Australian and Québécois-French speakers are considered "out-group" for Canadian-English speakers in terms of their linguistic background and nature of their accents. Our design does not preclude the possibility that in certain social contexts, Canadian-English and Québécois-French speakers may treat each other as "in-group" members (e.g., when working together on the international stage).

[2] Other factors may differentiate the two out-group accents to our listeners; for example, here the Québécois-French accent was more familiar to native Canadian-English listeners than the Australian accent. Our study did not aim to distinguish how these factors may have influenced the neural mechanisms underlying different out-group voices.

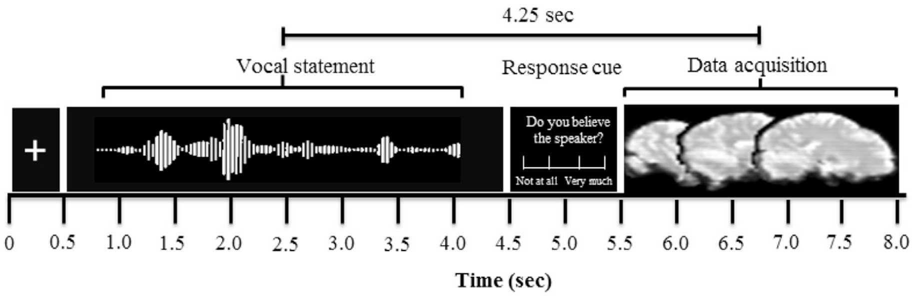

Fig. 1. Experimental design and task procedure. The timing of each event in one trial during the online fMRI task is presented. The onsets of the vocal statements of varied lengths were jittered by keeping the midpoint of the vocal stimulus 4.25 s away from the midpoint of the data acquisition period.

had the same trial length as the vocal events, except no sound was played. The whole scanning session consisted of six runs with the same stimuli composition. Each run started with three null events to allow for stabilization of the magnetization and lasted for 7 min 14 s. The experiment was controlled by Presentation software (Neurobehavioral Systems, Inc.).

2.5. Post-scan tests: accent favorability and intelligibility ratings

After the scanning session, each participant completed tests to assess their favourability and perceptual sensitivity towards each accent. First, participants indicated how favorable they feel towards Canadian-English, Australian, and Québécois-French speakers using a conventional explicit attitudinal measure (Greenwald et al., 1998). This measure was used to assess a listener's favourability when the explicit accent label was provided. On an illustration of a thermometer with a scale from 0 to 99, listeners rendered their evaluation by marking an appropriate position (where 1–49 = cold or unfavorable towards the speaker, 50 = neutral, and 51–99 = warm or favorable). Second, participants heard a subset of expressions from the fMRI experiment (12 per accent type) and judged how intelligible the speaker sounds on a 5-pt scale (from least to most).

2.6. fMRI data analysis

The fMRI data were analyzed using tools from the Functional Magnetic Resonance Imaging of the Brain software library (FSL, Smith et al., 2004). The following preprocessing steps were applied to all scans for each participant across runs: brain extraction (Smith et al., 2004); motion correction using MCFLIRT (Jenkinson and Smith, 2001); spatial smoothing with a 5-mm FWHM isotropic Gaussian kernel; and high-pass temporal filtering at 1/100 Hz. The first three volumes were removed from each run. Each participant's fMRI scan was then linearly registered to their corresponding T1-weighted structural MRI using a six-parameter rigid-body transformation. This was followed by nonlinear registration to the Montreal Neurological Institute (MNI)-152 standard space (Mazziotta et al., 2001). These were combined to transform all functional images into the MNI152 standard space. One participant was removed from further analysis due to unacceptable registration quality to the standard space.

A whole-brain general linear model (GLM)-based statistical analysis of the blood-oxygen-level- dependent signal (BOLD) was performed on a voxel-wise basis (Worsley and Friston, 1995). The first level analysis used FSL FILM with local autocorrelation correction for GLM time series analysis (Woolrich et al., 2001). Multiple models were built to assess the change of BOLD signal as a function of the speaker group, tone of voice, and participant's believability response. Each vocal stimulus was labeled as 'Native', 'Regional' or 'Foreign' and 'Confident', 'Doubtful' and 'Neutral' and modeled as durational events.

The first model focused on the main effect of accent, which involved contrasts between in-group and out-group accents while levels of confidence were collapsed, and the counterpart contrasts between out-group and in-group accents. The second model assessed the main effect of confidence in both directions while the speaker accents were collapsed (see Results section). The third model examined the interaction between confidence and accent (see Results section). The fourth model regressed the BOLD signal against participants' actual believability rating for in-group and out-group accents. The believability response was modeled from one to four, where one indicates "the speaker does not sound believable at all" and four indicates "the speaker sounds very believable". Participants' responses were modeled as an impulse response at the onset of the vocal stimuli (Jiang2017).[3] To assess whether common brain regions were active between accents, we used conjunction inference (Nichols et al., 2005) in reference to the uncorrected whole brain statistical maps from corresponding models of different accents.

2.7. Psycho-physiological interaction (PPI) analysis

A psycho-physiological interaction (PPI) analysis was performed to identify the synchronized neural activities between regions that permits a believability evaluation. The listener's believability response inference was predicted by the vocally-expressed confidence of the speaker (see below and Jiang et al., 2017). We assessed if the functional connectivity involving regional activity that are modulated by the speaker confidence is correlated with the believability judgement (O'Reilly et al., 2012). Consistent with previous reports on decoding vocal expression decoding (Frühholz and Grandjean, 2013; Kotz et al., 2013; Sammler et al., 2015; Schirmer and Kotz, 2006), our analysis identified the right superior temporal gyrus (STG) which was significantly correlated with the ability to rate vocal confidence in all accents, and other regions (including right IFG, right middle temporal gyrus (MTG), left SMA) that were correlated with the speaker confidence in each specific accent (see Supplementary materials). We searched the whole brain for any regions that are predicted to show synchronized activity to these regions. The PPI analysis also aimed to assess whether the functional connectivity was modulated by accent. Compared with the in-group accent, increased connectivity of

[3] Both positive and negative correlations were tested. Interaction contrasts were assessed ('Native positive correlation vs. Regional positive correlation', 'Native positive correlation' vs. Foreign positive correlation', 'Native negative correlation' vs. 'Regional negative correlation', 'Native negative correlation' vs. 'Foreign negative correlation'). Additional models regressing speaker believability with the BOLD signal in each accent were evaluated. For all models, a binary variable indicating a null event was included as a covariate. Each variable of interest was convolved with the double-gamma hemodynamic response function (HRF). Since all participants had multiple sessions with similar stimuli composition, a second level analysis was performed to combine whole brain statistical maps from the first level GLM time series analysis for each participant. This utilized FSL FMRIB's Local Analysis of Mixed Effects (FLAME) to perform fixed effects modeling (Beckmann et al., 2003; Woolrich et al., 2004). Final group level analysis used FLAME to perform mixed-effects analysis with automatic outlier deweighting to capture the mean group effect in both positive and negative directions, as well as the effect of participants' favorability or intelligibility of an accent on the underlying BOLD signals.

the right STG would be expected for out-group (Regional or Foreign) accents if the listener conducts a more detailed analysis of vocal confidence cues for evaluating the believability of out-group speakers.

To perform the PPI analysis, the right STG and other significant regions that were correlated with external confidence ratings were chosen as seed regions (Supplementary Material). These regions were defined by 9-mm diameter sphere centered at the peak activations in each of the seed region. The physiological activity (i.e. time series BOLD signal) from each region was extracted, and the BOLD signal correlation with the participant's believability response was considered psychological regressor. To evaluate the interaction between accent and believability response, we defined the regressor as the differential BOLD signals of believability response between native and one out-group accent. Two models were built with both including the native accent and each including one from the two out-group accents (Native vs. Regional or Native vs. Foreign). We also performed PPIs in each accent group by defining the regressor as the BOLD signals of the believability response in that accent. The first level analysis found the interaction between the underlying physiological activity and psychological regressor. Similar to the whole brain GLM analysis (described above), the second level analysis combined the statistical maps across runs for each participant, while the third level analysis captured the mean group effect, as well as the intelligibility perception and favourability towards a certain accent. We assess the effects of individual differences for each accent separately given that the related measures were obtained on each accent and the aggregate values would reduce the variation between listeners. For all whole brain GLM, PPI analyses and conjunction inference, areas of significant activation were identified using cluster thresholding with a Z cut-off of 1.96. All models were corrected for multiple comparisons at $p < .05$ using Gaussian random field theory (Worsley et al., 2002).

3. Results

3.1. Behavioral results: rating the believability of speakers with in-group and out-group accents

Fig. 2 displays the effect of accent on believability judgements for each of the three vocal expressions. Linear mixed-effects models included confidence level and accent as fixed effects, vocal expression and participant as random effects, and listener's gender, age and years of education as control variables. Believability ratings varied as a function of Confidence ($F(2, 6064) = 667.03$, $p < .0001$), Speaker Group ($F(2, 6064) = 7.96$, $p = .0004$), and the interaction between these two factors

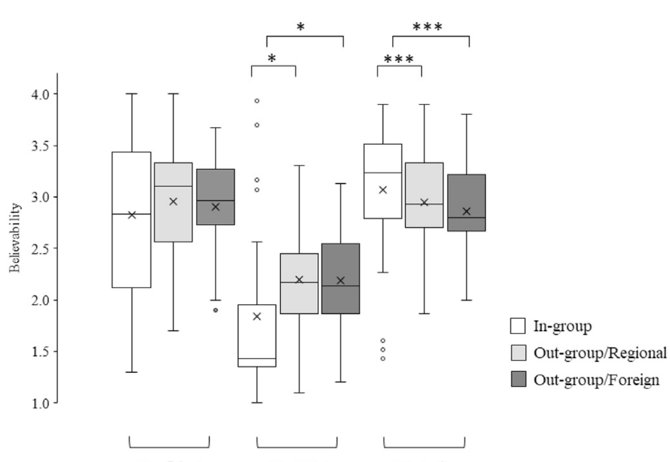

Fig. 2. Box and Whisker plots showing the online believability score for each accent and vocal expression type. The x in the box represents the mean score. The median divides the box into the range between the 1st and 3rd quartiles. Note: * 0.01 < p < .05; ***p < .001.

($F(4, 6064) = 20.53$, $p < .0001$). Overall, neutral and confident expressions were judged as more believable than doubtful expressions (neutral > doubtful: bs > 0.70, ts > 15.61, $ps < .0001$; confident > doubtful: bs > 0.74, ts > 16.49, ps < .0001), and listeners judged statements in their native accent to be more believable than those produced in out-group accents (Native > Regional: b = 1.45, t = 3.01, $p = .003$; Native > Foreign: b = 0.19, t = 4.02, $p < .0001$).

In the native accent, neutral statements were rated as more believable than confident expressions (neutral > confident: b = 0.22, t = 4.24, $p < .0001$), whereas believability ratings of neutral/confident expressions did not significantly differ in accented speech ($ps > .39$). Across speaker groups, neutral statements were more believable when produced in in-group than out-group accents (Native > Regional: b = 0.15, t = 3.78, $p = .0002$; Native > Foreign: b = 0.20, t = 4.88, $p < .0001$), whereas confident statements yielded similar believability impressions irrespective of group type ($ps > .47$). Doubtful expressions were *less* believable when produced in in-group than out-group accents (Native < Regional: b = −0.36, t = −2.17, $p = .03$; Native < Foreign: b = −0.36, t = −2.16, $p = .03$).

To contextualize patterns in the behavioral data, believability ratings were regressed in separate models on explicit ratings of speaker confidence for the stimuli (taken from the independent Pre-test group) and on ratings of accent favourability and intelligibility (taken from the Post-fMRI tests). There was a significant effect of confidence ratings on the believability judgment in all accents (bs > 0.42, ts > 19.62, $ps < .0001$). Regardless of speaker group, statements judged to be more confident were also considered to be more believable. A model revealed a positive effect of favourability towards the Québécois-French (Foreign) accent on the believability judgment (b = .008, t = 2.29, $p = .03$); those who were more favourable towards the Foreign accent label rated vocal expressions in that accent to be more believable. There was no evidence that intelligibility ratings influenced the believability judgment in any accent conditions ($ps > .23$).

3.2. fMRI results

When inferring believability about in-group speakers, Jiang et al. (2017) documented frontal– temporal networks responding to different levels of vocal confidence, with the left superior and inferior frontal gyrus more activated for confident statements, the right superior temporal gyrus for unconfident expressions, and bilateral cerebellum for statements in a neutral ("prosodically-unmarked") voice. Relative to in-group speakers, judging speaker believability from confident, doubtful, and neutral vocal expressions in an *out-group* accent yielded distinct patterns of cerebral activity, as reported in Table 2 and Figs. 3–4.

3.2.1. Main effects

For speaker group, the contrast "In-group vs. out-group/(Foreign + Regional)" revealed activations in the medial SFG, ACC, bilateral SFG, left middle occipital gyrus (MOG) and right MTG (Fig. 3a). The contrast "In-group vs. out-group/Regional" revealed activations in the ACC, medial SFG and ventral medial prefrontal cortex (vmPFC). The contrast "out-group/(Foreign + Regional) vs. Ingroup" revealed activations in the bilateral STG and MTG (Fig. 3b), whereas the contrast "out-group/Foreign vs. Ingroup" revealed activations in the left middle frontal gyrus (MFG), right IFG (triangularis), bilateral precentral gyrus (PreCG), left cerebellum, left superior occipital gyrus (SOG).

For the level of confidence, the contrast "Confident vs. Doubtful" revealed activations in the bilateral hippocampus and left fusiform (Fig. 3c). The contrast "Doubtful vs. Neutral" revealed activations in the right STG and left MTG (Fig. 3d). No activations were observed in any other contrast.

3.2.2. Interactions between speaker group and confidence

Given that the combination of out-group accents was heterogeneous and the accents are different in terms of their familiarity towards the

Table 2
Peak activations for functional contrasts involving different vocal expressions.

Brain Region	Number of Voxels	Z score	x	y	z
Main Effects of Expressed Confidence					
Confident vs. Doubtful					
L. Fusiform G.	771	3.26	−32	−38	16
R. Parahippocampal G.	677	3.08	24	−34	−12
Doubtful vs. Neutral					
R. Superior Temporal G.	582	3.12	46	−24	0
L. Middle Temporal G.	566	3.53	−64	−26	2
Main Effects of Accent					
In-group vs. Out-group (Regional + Foreign)					
L. Superior Frontal G.	2826	3.76	−14	62	20
L. Middle Occipital G.	2406	3.81	−34	−76	12
R. Middle Temporal G.	1913	3.51	40	−62	14
R. SMA	1273	3.29	0	−22	52
Out-group (Regional + Foreign) vs. In-group					
R. Superior Temporal G.	1808	4.09	50	−18	0
L. Middle Temporal G.	1361	4.03	−50	−22	2
Interaction between Expressed Confidence and Accent					
Out-group/Regional (Confident > Neutral) vs. In-group (Confident > Neutral)					
R. Caudate	7336	4.34	18	−4	26
Out-group/Regional (Confident > Doubtful) vs. In-group (Confident > Doubtful)					
R Rolandic Operculum	441	3.16	12	18	48
Out-group/Foreign (Confident > Neutral) vs. In-group (Confident > Neutral)					
L. Cuneus G.	5700	3.62	−6	−86	26
R. Fusiform G.	546	2.98	29	−45	−12
Out-group/Foreign (Confident > Doubtful) vs. In-group (Confident > Doubtful)					
R. Superior Parietal L.	2589	3.58	38	−54	58
R. Middle Temporal G.	1341	3.82	64	−42	6
Out-group/Foreign (Doubtful > Neutral) vs. In-group (Doubtful > Neutral)					
L. Cerebellum	546	3.69	−16	−48	−48
In-group (Confidence > Neutral) vs. Out-group/Regional (Confident > Neutral)					
Left Medial Superior Frontal G.	714	3.23	−8	38	34

listener and perceived intelligibility (see below), we first assessed the simple effects of speaker confidence for each out-group accent and for a dataset when the two out-group accents were combined. In the former analysis, the specific location of the neural activations were different between accents when we assessed the simple effects of speaker confidence in each out-group accent; whereas the latter analysis did not yield to any significant results (Supplementary Materials, S2). We then assessed the interaction models each time we included the in-group accent and one out-group accent, to reduce the impact of variability in the out-group accents on the statistical power of the interaction analysis.

The interaction analysis of the contrast ('Regional-Confident' vs. 'Regional-Neutral') > ('Native-Confident' vs. 'Native-Neutral') showed significant activation in the right caudate. While activity was increased in confident over neutral expression for the out-group/Regional accent, the opposite pattern was observed for the In-group accent (Fig. 4a). The interaction analysis ('Foreign-Confident' vs. 'Foreign-Neutral') > ('Native-Confident' vs. 'Native-Neutral') showed significantly greater activity in the left cuneus and right fusiform regions. While the out-group/Foreign accent exhibited increased activity in these regions for confident over neutral expressions, the In-group accent yielded increased response to neutral expressions (Fig. 4c and d).

The interaction analysis ('Regional-Confident' vs. 'Regional-Doubtful') > ('Native-Confident' vs. 'Native-Doubtful') revealed significant activation in the right rolandic operculum. While the out-group/Regional accent showed greater activity for confident over doubtful expression,

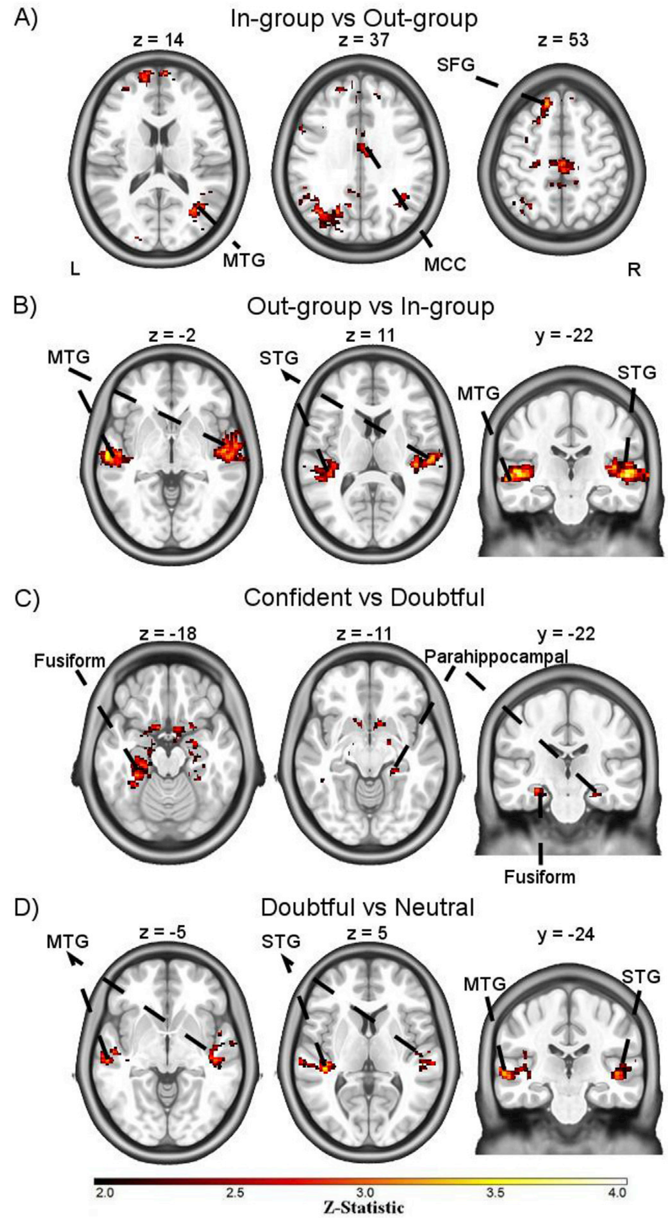

Fig. 3. Activation maps showing the difference in contrasts between in-group and out-group speakers regardless of vocal expression type: (a) In-group versus out-group (Regional + Foreign) accent; (b) out-group (Regional + Foreign) versus In-group accent; (c) Confident versus Doubtful voice; (d) Doubtful versus Neutral voice. All activations survived the threshold at cluster-level z > 1.96, $p < .05$ (GRF corrected).

this activation pattern was reversed in Native accent (Fig. 4b). The interaction analysis for this contrast ('Foreign-Confident' vs. 'Foreign-Doubtful') > ('Native-Confident' vs. 'Native-Doubtful') showed significant BOLD increases in right MTG and SPL for out-group/Foreign but not for In-group speakers (Fig. 4e and f).

The interaction analysis ('Foreign-Doubtful' vs. 'Foreign-Neutral') > ('Native-Doubtful' vs. 'Native-Neutral') showed significant BOLD increases in left cerebellum.

3.2.3. Parametric effects of perceived believability across speaker groups

When inferring believability from native accents, Jiang et al. (2017) reported BOLD signal changes in the SPL/PoCG: the right SPL was activated to a greater extent when in-group speakers were judged to be more believable. Here, when listeners judged speakers with *out-group* accents

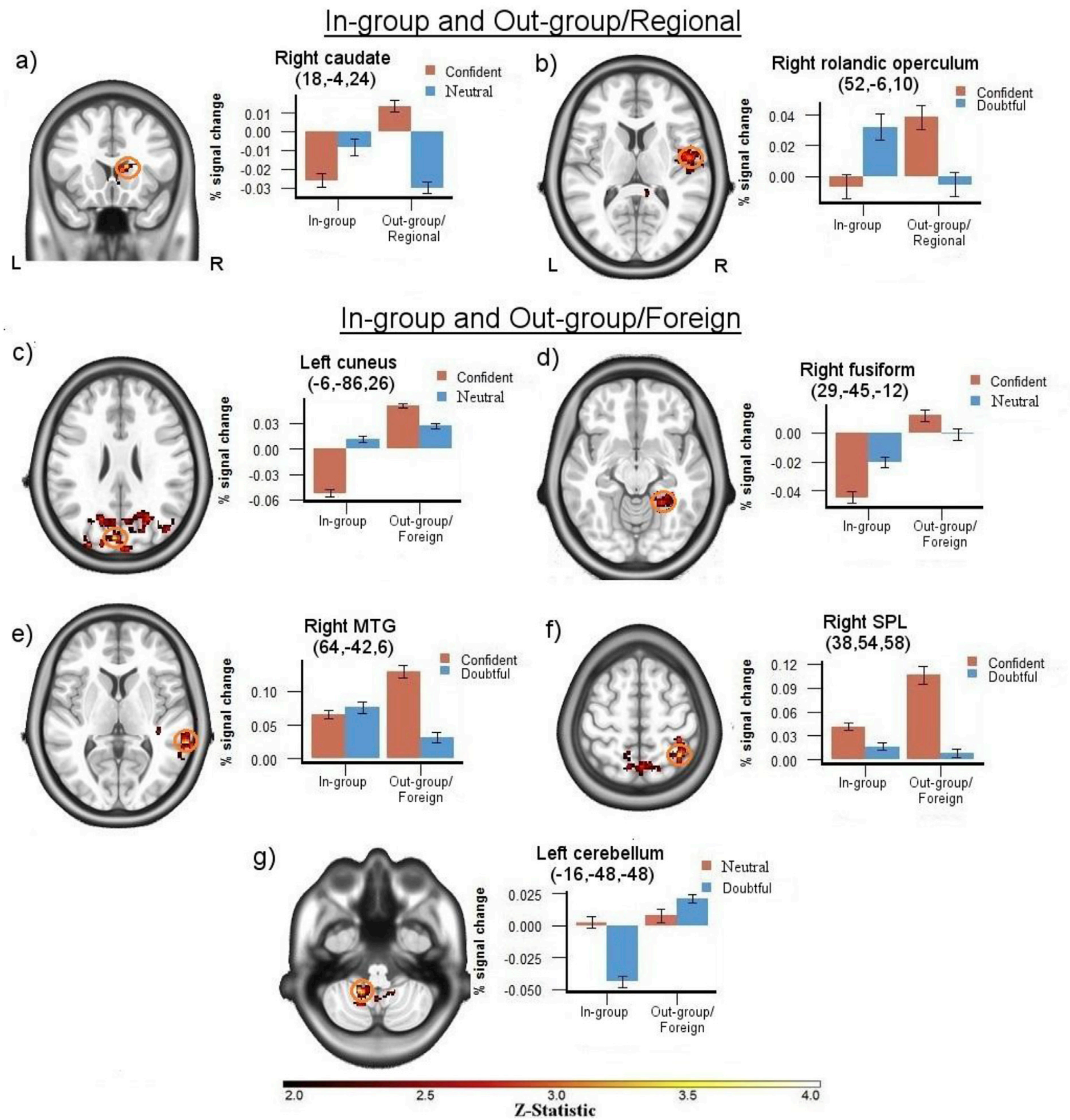

Fig. 4. Activation maps of contrasts showing the interaction of speaker accent ((a)–(b) In-group and out-group/Regional accent in the upper panels, and (c)–(g) In-group and out-group/Foreign accent in the bottom panels) and the vocal expression type. Bar graphs show relative differences in activation between vocal expression types for each accent. All activations survived the threshold at cluster-level $z > 1.96$, $p < .05$ (GRF corrected).

Table 3
Peak activations for parametric main effects of believability rating for out-group accents.

Parametric Effects of Increasing Believability Rating*					
Brain Region	Number of Voxels	Z score	x	y	z
Out-group/Regional accent					
L. Middle Temporal G.	2111	3.92	−58	−18	−2
R. Superior Temporal G.	1217	3.64	46	−14	2
L. Inferior Frontal G.	638	3.31	−54	26	16
Out-group/Foreign accent					
R. Superior Temporal G.	1678	3.85	58	−18	−2
L. Superior Temporal G.	1511	3.44	−38	−26	6

Notes: *Please refer to the results of the in-group accent in Jiang et al. (2017).

to be more believable, we observed increased activation in the bilateral STG, MTG, and heschl gyrus for both out-group accents; the left IFG was also activated for the out-group (Regional) accent in this condition (Table 3; Fig. 5a and b). Conjunction inference with the two out-group accents revealed common activations in the bilateral STG, MTG, and heschl gyrus (Fig. 5c). When the conjunction inference included all three accents, no common activations were significant.

3.2.4. Functional connectivity

Table 4 displays brain regions functionally connected to their corresponding seed regions by speaker group following the PPI analysis. The seed regions were selected as those which demonstrated a correlation with the independent speaker confidence rating (Supplementary Materials, Table S4). When the right STG was used as the seed region (Fig. 6a)

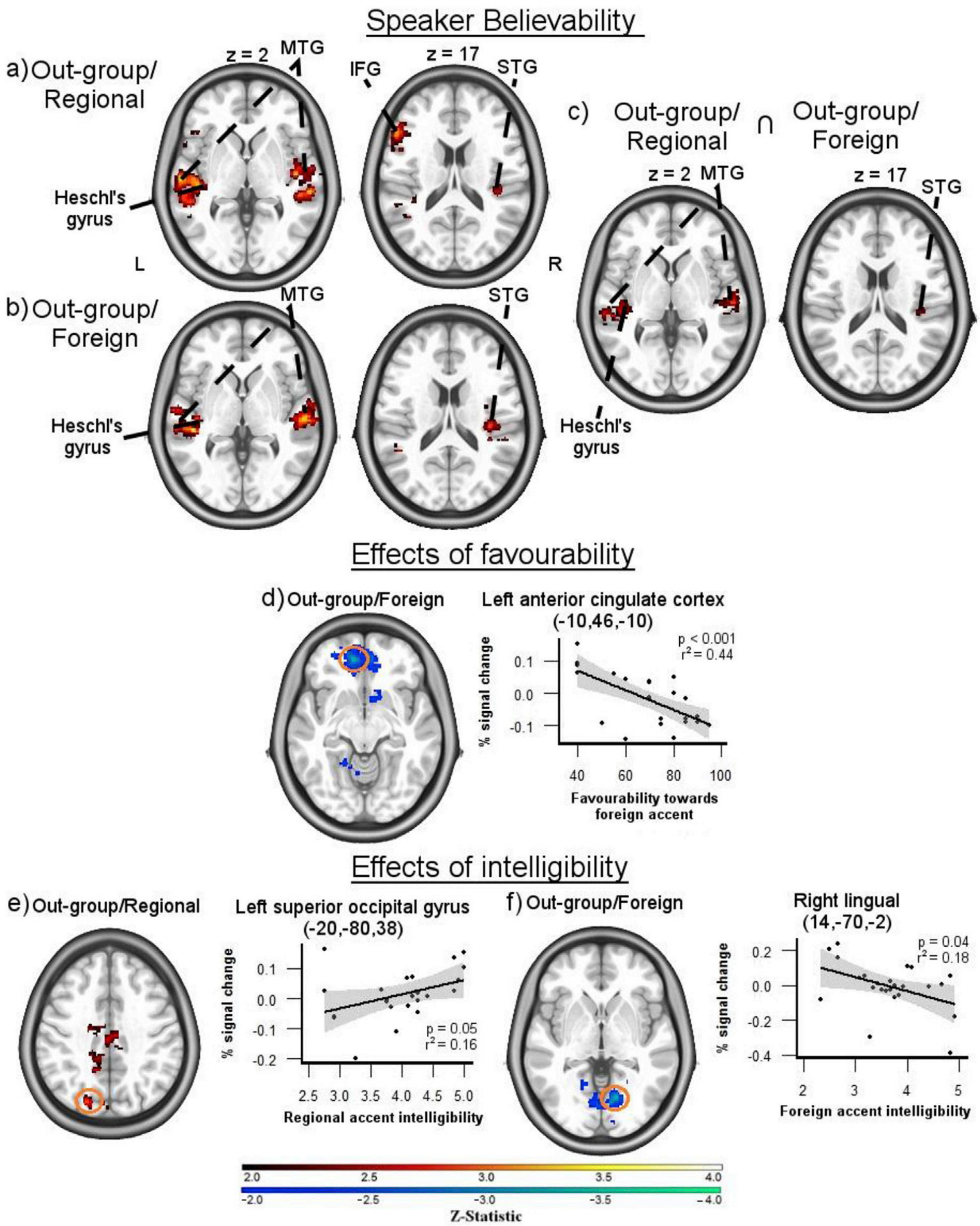

Fig. 5. Activation maps showing the regions surviving parametric analysis of believability rating for (a) out-group/Regional accent, (b) out-group/Foreign accent, and (c) the conjunction of the two out-group accents. Activation maps of (d) regions surviving the parametric analysis of increased believability in the out-group/Foreign accent, which was negatively modulated by accent favorability score; (e) regions surviving the parametric analysis of increased believability in the out-group/Regional accent, which was positively modulated by the intelligibility score; (f) regions surviving the parametric analysis of increased believability in out-group/Foreign accent, which was negatively modulated by the intelligibility score. All activations survived the threshold at cluster-level $z > 1.96$, $p < .05$ (GRF corrected). The percentage signal change was extracted from regions showing the maximal activation in the relevant cluster. The descriptive scatterplots are provided to complement the parametric maps, plotting the believability parametric effect on each individual sorted according to favorability and intelligibility ratings.

and participants' believability response was treated as the psychological regressor, significant correlations were found with the left inferior occipital gyrus for in-group accent, right ACC and MOG for out-group/Regional accent, and right supramarginal gyrus extending into PoCG for the out-group/Foreign accent. An interaction between speaker accent and believability ratings pointed to different functional connectivity with the right STG when listening to out-group vs. in-group speakers. When judging that a Foreign-accented statement was believable, the connectivity strength between the right MTG and right STG was significantly greater than for in-group speakers (Fig. 6b). When statements produced in a Regional accent were considered believable, the left lingual gyrus/MOG was significantly more connected to the right STG when compared to in-group accent (Fig. 6c).

Connectivity with other seed regions (right IFG, right MTG, and left SMA) was seen in each accent when participants judged speakers to be more believable (Table 4). For the in-group accent, the right IFG and left MOG were significantly correlated with increased believability, whereas correlations for the out-group/Foreign accent were found between right MTG and a set of regions including bilateral vmPFC, right medial SFG and ACC. Utterances rated as more believable in the out-group/Regional accent had greater connectivity strength between the left SMA and a set of regions including right PreCG, left PoCG, bilateral insular (inferior) and left IFG (triangularis, operculum). Also, greater connectivity strength was observed between the right IFG and the regions including bilateral MOG and bilateral fusiform.

3.2.5. Post-scan tests: individual differences in accent favourability and intelligibility

For the accent favourability ratings, participants rated in-group speakers as more favorable ($M = 86.28$, $SD = 16.62$) than out-group speakers (Regional: $M = 78.38$, $SD = 20.65$; Foreign speakers: $M = 67.80$, $SD = 20.06$, $ps < .0001$). Analyses showed that individuals who had a less favourable attitude towards the Québécois-French accent displayed increased responses in the vmPFC, ACC, middle cingulate cortex (MCC) and precuneus whenever they judged that the foreign-accented speakers were believable (Fig. 5d). No significant

Table 4
Peak activations for PPI analysis showing the parametric effects of believability rating. Seed regions were selected as the significant peaks in the parametric analysis of speaker confidence ratings.

PPI: Parametric Effects of Believability Rating[1]					
Brain Region	Number of Voxels	Z score	x	y	z
IIn-group accent: Increasing believability rating; right IFG as seed					
L. Inferior Occipital G.	654	3.22	−24	−96	−6
Out-group/Regional accent: Increasing believability; left SMA as seed					
L. Middle Frontal G.	1198	3.56	−34	42	28
R. Inferior Frontal operculum	923	3.56	52	14	2
L. Insula G.	807	3.4	−30	12	−6
L. Middle Temporal G.	555	3.21	−56	−58	0
Out-group/Regional accent: Increasing believability; right IFG as seed					
Left Middle Occipital G.	3220	3.39	−22	−92	2
Out-group/Foreign accent: Increasing believability rating; right MTG as seed					
R. ACC	1453	3.27	8	46	8

PPI: Interaction between Believability Rating and Accent					
Brain Region	Number of Voxels	Z score	x	y	z
(Out-group/Regional > IIn-group accent) × Increasing believability; right STG as seed					
L. lingual G.	780	3.17	−16	−74	−4
(Out-group/Foreign > IIn-group accent) × Increasing believability; right STG as seed					
R. Middle Temporal G.	689	3.48	38	−58	16

associations between favourability and believability ratings were noted for the other accents. For intelligibility ratings, the native accent was judged as more intelligible overall (M = 4.24, SD = 0.92) than the Regional (M = 3.97, SD = 0.77) and Foreign (M = 3.71, SD = 0.77, *ps* < .0001) accents. There was a positive relationship between increased speaker intelligibility and increased believability, yielding increased responses in the precuneus (in-group accent) or in the cuneus, left MOG, and MCC (out-group/Regional accent). For the Foreign accent, there was a negative relationship between speaker intelligibility and believability; when statements by less intelligible speakers were judged to be believable, regional activity increased in the left cerebellum and lingual gyrus. These findings show that subjective impressions of speaker favourability and intelligibility modulate the believability-associated neural response as a function of accent type (Fig. 5d–g).

4. Discussion

The present fMRI study investigated how social categorization of a speaker's accent could modulate the functional activity that underlies inferences about whether to believe statements that vary in vocally-expressed confidence. The behavioral ratings showed that speaker believability increases as a function of the level of *vocal confidence* expressed by both out-group and in-group speakers (Jiang et al., 2017). Independent of accent type, we found increased neural activity in the right superior and left middle temporal regions in doubtful vs. neutral expressions, which furthers our knowledge of how the brain decodes vocal uncertainty from in-group speakers to a more diverse dataset with varied speaker groups (Jiang et al., 2017). This result is consistent with previous literature where the right STG appears to be involved in decoding sentence-level suprasegmental information (Grandjean et al., 2005; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006) while the left MTG likely participates in extracting sentence-level semantic information (Friederici, 2002; Vandenberghe et al., 2006). Given evidence that doubtful voices possess salient acoustic cues that indicate a speaker's hesitation or feeling of unknowing (Jiang and Pell, 2017), our data suggest that when judging believability, vocally-expressed doubt may require greater demands in auditory cortical areas involved in extracting the emotive value of suprasegmental speech cues irrespective of speaker identity.

Similarly, the increased activity we observed in bilateral hippocampi in confident versus doubtful statements may be linked to selective processes that *re-encode an event* that is confirmed or highly-known (Barbeau et al., 2005; Brown and Aggleton, 2001; Maril et al., 2003; Maril et al., 2005). Here, the confident voice may function as a confirmation that facilitates the re-encoding of a highly-known statement (e.g. She has access to the building), a condition in our study that relied selectively on functions of the hippocampus. Given the novelty of this observation, inferring the mental process underlying these neural activations has to be further validated with large-scale decoding methods, such as multivariate pattern recognition, that have the capacity to formally test the ability to infer mental states from neuroimaging data (Poldrack, 2006, 2011). In contrast to Jiang et al. (2017), the fact that our confident voice did not reveal increased activity in a fronto-temporal network involved in prosody perception (e.g., Sammler et al., 2015) when the accents were combined may be due to increased variability introduced by the different out-group accents for this specific vocal expression (see below). Future designs that elaborate on the neural correlates of speaker confidence processing for different voice types, as well as in the context of general knowledge statements that possess an inherent truth value, will be

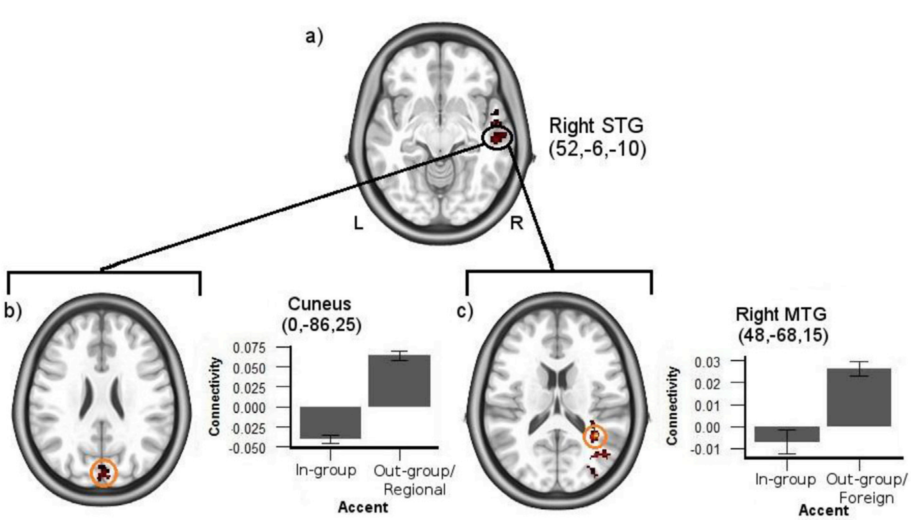

Fig. 6. Activation maps showing the right STG (a) that was shown in the conjunction analysis across all accents in the parametric analysis of the independent confidence ratings. Activation maps for regions that were more strongly connected to the right STG in the parametric analysis of believability rating for the accented speakers ((b) out-group/Regional; (c) out-group/Foreign) than for the In-group speaker. Bar graphs showing the level of activity that was associated with the increased believability rating in each accent.

immensely valuable.

4.1. Effects of out-group voices on vocal confidence and believability

In response to our first two main questions, we found that in-group voices were generally judged to be more believable than speakers with out-group (regional or foreign) accents. Moreover, forming believability impressions were associated with distinct neural activity when out-group accents were present. When broadly construed, our results suggest that listeners categorized speaker group membership based on their accents, and that implicit evaluation of out-group voices was associated with negative social repercussions (Bradlow et al., 1999; Nygaard and Pisoni, 1998). The results also exemplify that human vocal characteristics modulate perceptual experiences and socio-emotional reactions towards a person who carries an out-group accent (Bestelmeyer et al., 2015; Hayakawa et al., 2016; Van Berkum et al., 2008; Van den Brink et al., 2012).

Closer inspection of the behavioral data showed that the impact of out-group accents on believability impressions varied according to characteristics of a speaker's vocal expression. Neutral expressions, which do not provide explicit vocal cues about a speaker's confidence level, were rated as more believable when produced by in-group speakers compared to out-group speakers. In contrast, statements produced in an overtly confident tone were rated as equally believable across accent types and doubtful statements were judged *less* believable when expressed by in-group speakers (Fig. 2). The differential modulation of speaker accent on believability judgments for neutral and confident voices implies that early processes for social categorization may promote differential use of vocal stimulus features to make inferences about the out-group speakers. Supplementary analyses revealed that while listeners relied on acoustic cues, such as utterance duration to make believability inferences regardless of speaker group, they relied on a wider range of acoustic parameters to infer believability from out-group speakers (e.g., mean and variation of f0 and mean intensity). This meant that believability decisions about out-group speakers were dependent to a larger extent on how listeners used acoustic information to derive the speaker's *confidence* level to guide believability ratings. These patterns suggest an important interplay between implicit evaluation processes that mark the out-group status of a speaker and those that act on stimulus-related acoustic features for assessing confidence and believability. We speculate that in the case when out-group speakers sounded confident, comparative procedures appeared to be engaged to counteract a listener's tendency to judge out-group accents as less believable or trustworthy (see also Lev-Ari and Keysar, 2010).

The conclusion that believability impressions formed about another person's voice differ according to the group status of the speaker was supported by the neuroimaging data. When out-group and in-group accents were compared, we noted increased temporal activity in response to the out-group speakers (bilateral STG and MTG). This supports the idea that listeners engaged in more extensive stimulus-related acoustic processing when socially evaluating individuals with an out-group accent (these points are elaborated below). When the main effect of in-group vs. out-group voices was tested, we observed greater activation of both ACC and vmPFC when listeners evaluated in-group speakers. This latter finding supports previous research that suggest these regions may have heightened empathy-related responses towards individuals from the same group (for example, when people viewed pictures describing in-group members suffering from physical pain, Chiao and Mathur, 2010; Mathur et al., 2012; also, Masten et al., 2011). Other research shows that mPFC activity was reduced when out-group members triggered negative feelings in a perceiver (e.g. disgust, Harris and Fiske, 2007). Here, modulation of the ACC and mPFC was elicited by the speaker's accent (especially when in-group speakers were compared to the out-group/regional accent), suggesting that the manner of pronunciation can trigger systematic responses related to group category knowledge, which may be associated with greater empathy towards in-group speakers, or "in-group favoritism" (Amodio, 2014; Han, 2018; Katsumi and Dolcos, 2017; Shkurko, 2013).

4.2. Interplay of accent and vocal confidence information

Our main theoretical focus was how in-group/out-group voices *interact* with vocal confidence information to form believability impressions. One way that speaker identity influenced this process can be seen in how the brain responded to confident voices in each condition. As noted earlier, participants rated confident voices to be equally believable across accent types; however, activity in the right MTG and SPL (out-group (Foreign accent) and right rolandic operculum (Outgroup/Regional accent) was greater for confident expressions produced by the out-group speakers. Perceptual data revealed that out-group voices were assigned slightly lower confidence ratings than the in-group voices for both confident and neutral expressions overall (Table 1); however, analyses confirmed that differences in the perceived strength of confidence between accents could not explain differences in the activation patterns for the out-group vs. in-group voices. There could be two possible interpretations for these temporo-parietal activations. First, these temporal and parietal regions may be part of the larger mirroring and mentalizing networks for rendering social attributions about other people (Van Overwalle and Baetens, 2009). For example, the mirroring network, has been shown to rapidly and intuitively read a speaker's nonverbal behavior and recognize the goal of a perceived action. This network was also found to be more activated in prosodically-unmarked expression in the in-group voice (Jiang et al., 2017). Meanwhile, the mentalizing network appears to be involved in the identification of speaker's mental state at an abstract semantic level, especially when the vocal cues are ambiguous (Van Overwalle and Baetens, 2009). The second interpretation is the right SPL and MTG may serve to extract higher-level meaning from lower-level acoustic cues (Hickok and Poeppel, 2007), as they were reported to be sensitive to *acoustic* differences underlying vocal confidence expressions when judging whether in-group speakers are believable (Jiang et al., 2017) and may therefore serve to extract higher-level meaning from lower-level acoustic cues (Bestelmeyer et al., 2012; Sammler et al., 2015). More investigations are warranted to disentangle these two possibilities. Given that regions associated with deriving social meanings from lower-level acoustic variations in vocal confidence expressions were relatively more engaged when speakers had an accent, results suggest that subsequent believability impressions were based on a more detailed analysis of vocally-expressed confidence in the context of accented vs. unaccented speech (Fiske and Neuberg, 1990; Perrachione et al., 2010; Jiang et al., Under Review). Specifically, vocal signals that intentionally mark the high confidence level of a speaker were the object of increased analysis when out-group voices were encountered, highlighting the interplay of these two sources of information when out-group speakers communicate believability and trust. This conclusion fits well with event-related potential data showing that early perceptual processes for registering the motivational salience of vocal confidence cues are categorically different when listening to in-group vs. out-group speakers. When speakers have an in-group accent, expressions of doubt are considered more salient for evaluating whether a statement is believable, whereas the opposite is observed for out-group accents, with increased brain responses for confident vocal cues for these speakers (Jiang et al., Under Review).

Another way that out-group voices may have affected the neural activity underlying believability decisions refers to a common set of regions that showed heightened activation for *different* vocal confidence expressions depending on the status of the speaker. Activity in the caudate (Regional accent) and left cuneus/right fusiform (Foreign accent) were driven by *confident* expressions produced in out-group accents, whereas activity in the same regions were modulated by *neutral* expressions in the native accent. Jiang et al. (2017) reported that the caudate facilitated believability inferences from the voice and were recruited to a greater extent when listeners actually believe a speaker, as well as in individuals who are more attuned to social relations. The caudate has been suggested to play a role in temporal prediction (the decoding of vocal behaviour via specific prediction of temporal patterns with the unfolding of speech) and potentiating meaning from emerging vocal representations (the derivation of speaker intention from nonliteral speech; Pell and Leonard, 2003; Pell et al., 2014; Schwartze and Kotz, 2013). For in-group speakers, it can be said that greater demands are placed on the caudate mechanism for predicting the socio-emotive value of stimuli that *lack* explicit vocal cues of confidence (*neutral* expressions), a mechanism that contributes more broadly to impressions of increased believability and trust (Jiang et al., 2017). However, this mechanism appears to be engaged more extensively by *confident*-sounding voices to potentiate believability decisions about out-group/regional speakers, possibly pointing to differential use of vocal confidence information in the context of in-group and out-group voices.

For in-group speakers, increased activation of medial temporo-occipital regions by neutral statements (Jiang et al., 2017) was ascribed to the *level of ambiguity* associated for construing the speaker's mental state when explicit vocal cues of confidence were lacking; here, similar regions were activated by confident expressions in the out-group/foreign accent. Interestingly, when confident statements were produced by out-group speakers, these expressions made a markedly weaker impression of confidence on Canadian listeners (especially in the foreign accent, Table 1). This implies that out-group statements produced in a confident manner were less specified in terms of the speaker's ‘feeling of knowing’, placing higher demands on processes for abstracting and rendering believability attributions about out-group speakers. It has been reported that confident voices receive more in-depth processing when out-group accents are detected because this situation foregrounds conflicting tendencies induced by social categorization (tendency to disbelieve an out-group speaker) and those triggered by stimulus-related features (tendency to believe a confident voice, Fiske and Neuberg, 1990; Freeman et al., 2010; Perrachione et al., 2010). For out-group speakers, this conflict is hypothetically resolved through a stronger analysis of intended motivational properties of their voice–i.e., a more extensive “piece-meal” analysis of vocal confidence cues (Fiske and Neuberg, 1990)–for the listener to reconcile whether or not to believe their statement.

By examining the interaction of speaker group and vocal expression type, we demonstrated that greater neural activations were always observed for particular vocal expressions that were perceived as most believable *in relation to the social group* (i.e., neutral statements for the in-group voice and confident utterances for the out-group voice). The idea that believability impressions about in-group and out-group speakers were arrived at using different processing mechanisms is further supported by analyses correlating neural activity with listeners' actual believability ratings, which revealed selective correlation in the superior temporal cortex only for out-group accents (see below).

4.3. Effects of out-group accents on functional networks for inferring believability

To answer our third question, we looked directly at brain activity associated with graded decisions about *how much* listeners actually believed a statement (irrespective of vocal expression type). Different functional networks were observed for in-group and out-group speakers. Previous work revealed that a parietal network involving the bilateral SPL may be responsible for this inferential process for in-group speakers (Jiang et al., 2017). These regions form part of the hypothesized mirroring network (Kuhlen et al., 2015; Van Overwalle, 2009) and were modulated by attentional saliency (Liao et al., 2010; Molenberghs et al., 2012; Wu et al., 2011) for mentalizing speaker or actor characteristics.

In contrast, rendering the same decision about an out-group speaker selectively recruited bilateral superior and middle temporal gyri, highlighting a functional dissociation in how representations of believability are constructed for in-group speakers (bilateral SPL) and speakers with an accent (bilateral STG). The bilateral temporal activation observed in superior, middle, and Heschl's gyri, were associated with increased believability ratings about out-group speakers may form part of the ventral stream of speech perception that starts with spectrotemporal analysis in the bilateral dorsal STG, phonological processing from the middle to posterior STS; followed by integration of lexical and contextual information in the MTG (Hickok and Poeppel, 2007). Earlier work on accent processing (Adank et al., 2012) reported a similar temporal/parietal dissociation, with enhanced temporal responses to unfamiliar vs. native accents (in the left posterior STG) and enhanced parietal responses for native vs. unfamiliar accents (right supramarginal + angular gyrus). While our data extends evidence of this dissociation to the evaluation of speaker believability, further work is required to verify these results.

PPI analyses explored the functional connectivity with the right STG, a region that may encode suprasegmental information as salient and acoustically complex socio-emotional events (Grandjean et al., 2005; Kotz et al., 2006; Schirmer and Kotz, 2006; Wildgruber et al., 2006). Local patterns of activity in the right STG were sensitive to different vocal expressions of basic emotion (Kotz et al., 2013). The anterior portion of the right STG/STS has been linked to speaker identity identification (Belin et al., 2002; Belin and Zatorre, 2003; Von Kriegstein and Giraud, 2004), whereas posterior STG responds to voice familiarity (Von Kriegstein and Giraud, 2004) and acoustic variations associated with nonverbal expressions (Bestelmeyer et al., 2011; Rauschecker and Scott, 2009). Our supplementary analyses demonstrate that both anterior and posterior portions of right STG are critical for judging the vocal confidence level of speakers from dynamic changes in acoustic variables, such as speech rate, irrespective of accent type. Responses in the right STG increased with impressions of higher speaker confidence for both in-group and out-group speakers, providing new evidence for the role of the right STG in the decoding of voice information that refers to a speaker's mental state (Jiang et al., 2017). Our evidence supports models of speech perception that advocate a critical role of the right superior temporal cortex in the integration of utterance-level prosodic cues, such as pitch, stress, syllabic boundary, and syllabic rate (Friederici and Alter, 2004; Hickok and Poeppel, 2007).

When listeners heard out-group accents, there was distinct functional connectivity between regions for decoding vocal confidence expressions (e.g. right STG; right MTG; right IFG) and regions that have been shown to be part of mentalizing and inferring meaning from speech acts (e.g. bilateral PoCG, Jiang et al., 2017; right ACC/medial SFG, left SMA, Egorova et al., 2017; Sammler et al., 2015; insula: Rigoulot et al., 2014; and left IFG: Cavallo et al., 2015; Feng et al., 2017). The pattern of functional connectivity in right STG was uniquely involved in the out-group accents relative to the in-group accent. When listeners made the decision that an out-group speaker was believable, the connectivity strength increased between the right STG and right MTG, extending to the MOG (Foreign accent), and between the right STG and lingual gyrus/MOG (Regional accent, Fig. 6), when compared to in-group speakers (see further discussion below). This connectivity was not evident when evaluating utterances spoken in the native accent. This could point to a phase shift in the BOLD signal of one region relative to the other when an in-group speaker was evaluated, and a synchronization of the signal when an out-group speaker was evaluated (Chen et al., 2011).

As the anterior STG – posterior MTG coupling may serve as a ventral processing stream for mapping phonological to semantic/conceptual representations (Hickok and Poeppel, 2007), greater activity in this pathway while evaluating foreign-accented speakers suggests that operations for integrating vocal confidence cues and their social meanings were taxed to a greater extent by the out-group/foreign accent (Hickok and Poeppel, 2007). In both out-group accents, increased connectivity between the right STG and MOG again underscores that vocal expressions, rather than personal characteristics of a speaker, were analyzed more extensively by listeners to decide whether they should believe out-group vs. in-group speakers. Interestingly, in a separate analysis, activity in the lingual gyrus for Foreign accent and cuneus and left MOG for Regional accent was related to how *intelligible* out-group speakers were. The perceived intelligibility of out-group accents may constitute a general factor that increases processing demands when inferring speaker believability, modulating activity in medial occipital regions.

4.4. Individual differences

With respect to our final question about the role of individual differences, we found that the extent to which listeners perceived out-group speakers to be intelligible influenced the neural mechanisms underlying believability decisions, despite the fact that intelligibility did not seem to significantly impact believability outcomes (i.e. behavioral ratings). Intelligibility can affect speaker perception in diverse ways; Bresnahan et al. (2002) demonstrated a positive relationship between intelligibility of an out-group accent and one's (positive/negative) attitude towards the accent, but this depends on individual listener characteristics (e.g. one's level of commitment to an ethnic group). Given that the intelligibility rating we gathered was based on a selective subset ($n = 12$) of vocal expressions produced by two speakers, we refrain from drawing major conclusions about the importance of accent intelligibility on believability inferences due to our small data sample. Nevertheless, our findings support the hypothesis that intelligibility differentiates native and unfamiliar accents in the medial occipital cortex, the premotor and motor regions (Ackermann and Rieker, 2004; Ghazi-Saidi et al., 2015), *and* that accent could mediate how we derive social inferences from the voice by differentially engaging these regions.

We also examined how listener attitudes affected evaluations of a given speaker group, by measuring each a listener's subjective "favourability" towards an accent when confronted with its corresponding label in post-fMRI tests. In listeners who reported feeling *less* favorable towards the Québécois-French accent, neural responses to statements judged to be believable elicited greater activation in cortical midline structures, including bilateral ACC and precuneus. As mentioned earlier, these regions may be part of the mentalizing network normally recruited when individuals empathize with members of their own group (see Chiao et al., 2013 for a review; Han et al., 2017; Luo et al., 2015). These regions may also regulate conflicts between social expressions and stereotypical information accessed from the target's identity; as shown by Hehman et al. (2014), the ACC and mPFC are activated more when study participants view an out-group member displaying a stereotypically incompatible facial expression. Along the same lines, there was a stronger connectivity strength between left PoCG and bilateral ACC (extending to left vmPFC) in individuals who did not tend to trust others in their daily lives when listening to statements that they actually consider to be believable (Jiang et al., 2017).

These results suggest that when listeners hold a negative social bias (e.g., towards a particular race or accent, a tendency not to trust), there is increased recruitment in areas underlying the monitoring of socially-conflicting information to arrive at a *positive* evaluation of the speaker (here, to judge that a particular statement is believable). These results are supported by the literature that demonstrate increased favoritism of in-group characteristics and decreased favoritism of out-group members (Coupland and Bishop, 2007; Edwards, 1982; Hurt and Weaver, 1972; Mulac et al., 1974; Lev-Ari and Keysar, 2010; Ryan and Sebastian, 1980). Contrary to Bestelmeyer et al.'s (2015) study which used an adaptation paradigm, we did not observe activation in the amygdala for either in-group or out-group stimuli, an effect that Bestelmeyer et al. attributed to heightened emotional reaction to in-group accents which are more socially relevant to the listener. The fact that our task guided participants' attention towards specific cues that reveal a speaker's emotive or *mental* state in relation to their utterance (feeling of (un)knowing, Caffi and Janney, 1994; Jiang and Pell, 2016a; b), rather than passive listening to in-group and out-group voices, could account for differences in the amygdala response between studies, pending new work.

4.5. Towards a neurocognitive model of person perception from in-group and out-group voices

What neurocognitive architecture supports social inferences from the voice and how does this vary for in-group and out-group speakers? Based largely on the visual modality, theoretical and empirical work suggests two main mechanisms underlying person perception: the identification of category membership based on distinctive characteristics (skin colour, accent, etc.); and the individuation of personal attributes (confidence, trustworthiness, etc.). According to Fiske and Neuberg (1990), categorical perception functionally precedes an attempt to attribute personal characteristics during social perception, and there is evidence that social category information affects individuation processes, delaying how social attributes are derived by out-group perceivers (Freeman et al., 2010). Although the voice is rarely studied, it is clear here that accent-related features provide sufficient group category information to identify the out-group status of a speaker, and that this process may promote differences in the neural mechanisms used to make social attributions about individuals who do or do not share group membership with the listener.

The observed differences in regional brain activity and functional connectivity we report for out-group voices serve as a stepping stone to elaborate a neurocognitive model of person perception derived from different sources of voice information. Our results point to partly distinct processing routines for using voice information as a basis for social decisions about in-group vs. out-group speakers. In the case of in-group speakers, listeners may draw directly upon their personal experience and shared knowledge of "nonverbal accents" (Tesink et al., 2009; Van Berkum et al., 2008) to construe speaker attributes such as believability. Native speakers provide clues and encode believability-related traits in the voice that become familiar to the listener (trustworthiness, competence), learned through social convention or cultural norms (Oleszkiewicz et al., 2017; McAleer et al., 2014; Tsantani et al., 2016). According to some data, this processing is supported by increased involvement of mentalizing/mirroring networks for inferring person attributes based on in-group social signals (e.g. aIPS for voice in Jiang et al., 2017; mPFC for face in Freeman et al., 2010). We hypothesize that this socially-derived knowledge can often be employed in a heuristic, "top-down" manner for mentalizing the speaker meaning directly from the in-group voice, although additional neural resources may at times be (re)deployed in cases when in-group speakers do not clearly mark their vocal intentions (e.g., by recruiting basal ganglia and medial temporo-occipital regions when in-group speakers produce statements in a "neutral" tone, Jiang et al., 2017).

Our data suggest that out-group voices involve a less direct path for inferring social characteristics such as believability, one that draws upon a more detailed analysis of distinctive features supplied by the speaker. When faced with atypical or unexpected vocal characteristics that mark speakers as belonging to an out-group, listeners may accord greater weight to stimulus features for inferring personal attributes of an out-group speaker's voice. These stimulus features maybe used by the listener to reanalyze the vocal characteristics for the purpose of making social evaluations. In the present example, the listener tends to base a believability decision on the perceived level of vocally-expressed confidence, with more extensive recruitment of temporal voice areas (particularly right STG, MTG). When the speaker conveys an out-group accent, the perceived confidence rating mediated the relationship between the duration of utterance and believability, meaning that the longer the utterance, the lower the speaker confidence, which resulted in the lower believability rating (Supplementary Materials, S2). This suggests that the cognitive system may attempt to integrate and reconcile these features with knowledge of the (perceived) out-group to form a holistic believability impression. It is also possible that listeners engage in a more detailed analysis of vocal expressions in accented speech due to a basic shift in processing strategy linked to their perceived psychological distance from out-group speakers (Hayakawa et al., 2016; Jiang et al., In review). These ideas will foster the development of a more detailed model of how in-group and out-group voices affect social inferencing and person perception by comparing evidence from functional brain activity with neurophysiological data that pinpoint the underlying cognitive processing structure with greater precision (Jiang et al., In review).

4.6. Conclusion and future directions

Using an fMRI design that orthogonally manipulated speaker group (via accent) and vocal expressed confidence, we found that in-group accents were perceived as more believable than out-group accents, supporting the idea that a categorization process takes place based on a speaker's accent. Importantly, we observed that superior parietal and middle temporal regions, responsible for extracting speaker meaning from lower-level acoustic variations, were activated by confident-sounding voices only for the out-group speakers. Basal ganglia and temporal-occipital regions, necessary to abstract more ambiguous speaker meanings when in-group speakers produced statements in a neutral voice, were instead activated by the *confident* expressions produced by out-group speakers. Moreover, for the out-group speakers, stronger impressions of believability modulated activations in the bilateral superior and middle temporal regions, and enhanced functional connectivity between the right STG (parametrically associated with the perceived speaker confidence) and the medial temporo-occipital and right MTG. These findings suggest that a shift in neural processing mechanisms may occur based on a speaker's in-group status, with out-group speakers triggering a more extensive analysis of stimulus features (vocal cues of confidence) when social inferences are made.

The experimental setting in our study creates a minimal scenario to examine the perception of individuals with different mother tongues or cultural backgrounds, which can be generalized to many real-life situations in which speakers are treated as members of an out-group (e.g., when a physician with a foreign accent provides diagnostic information to a patient). Despite the fact that second language speakers involuntarily transfer phonetic variations of their first language to their second, whereas speakers of a regional accent simply display dialectal variations of their native language, our data uncovered few major differences in how the *type* of out-group accent affected neural regions for inferring speaker believability, extending our knowledge regarding the social consequences of speech accent and the neural mechanisms underlying person perception via multiple vocal/speech cues. Future designs may allow more focused comparisons of how different types of out-group accents influence social evaluative processes in the brain. The important interplay between speaker identity and vocal emotive processing, highlighted here and by previous psychophysical and neuroimaging evidence (Andics et al., 2010; Bestelmeyer et al., 2015; Campanella and Belin, 2007; Perrachione and Wong, 2007; Perrachione et al., 2010), merits further investigation.

While the status of our out-group accents was defined here in linguistic terms—accent information provides a unique and stable cue to inform a social category (Bestelmeyer et al., 2015; Bresnahan et al., 2002) and is weighed heavily to formulate social groups in daily communicative settings—it should be noted that the contexts that define the group relationship between the speaker and listener are malleable. This can alter when accented speakers are classified as in-group vs. out-group members to the listener. For instance, our Québécois-French speakers may be treated as 'in-group' members to English-Canadians in certain social settings, for example, that emphasize shared history or nationality. Thus, context affects the group relationship between the listener and the speaker, with important implications for neurocognitive processing. Comparing the modulatory effects of out-group accents when social groups are unambiguously defined based on linguistic vs. social criteria (e.g., based on political beliefs, Lau and Cikara, 2017) is a topic that merits future examination. Future studies should also examine how social learning processes, such as systematic exposure to a foreign language or dialect, modulate neural mechanisms underlying social inferences for out-group voices (Cao et al., 2015; Han and Ma, 2016; Liu et al., 2016; Orena et al., 2015; Perrachione and Wong, 2007; Zuo and Han, 2013).

Acknowledgment

We acknowledge all participants in the study. This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery grant, RGPIN/2016-04373) and a James McGill Professor award to M.D. Pell, and the McGill McLaughlin Scholarship to X. Jiang. We appreciate the valuable comments received at the Cognition and Circuits Seminar organized by Dr. Lesley K. Fellows at the Montréal Neurological Institute. We thank Ms. Kelly Hennegan and Dr. Pan Liu for their precious help in testing participants, and Michael Ferreira for his technical consultation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2018.07.042.

References

Ackermann, H., Rieker, A., 2004. The contribution of the insula to motor aspects of speech production: a review and a hypothesis. Brain Lang. 89, 320–328.
Adank, P., Evans, B., Stuart-Smith, J., Scotti, S., 2009. Comprehension of familiar and unfamiliar native accents under adverse listening conditions. J. Exp. Psychol. Hum. Percept. 35, 520–529.
Adank, P., Davis, M., Hagoort, P., 2012. Neural dissociation in processing noise and accent in spoken language comprehension. Neuropsychologia 50, 77–84.
Amodio, D., 2014. The neuroscience of prejudice and stereotyping. Nat. Rev. Neurosci. 15, 670–682.
Andics, A., McQueen, J., Petersson, K., Gál, V., Rudas, G., Vidnyánszky, Z., 2010. Neural mechanisms for voice recognition. Neuroimage 52, 1528–1540.
Bach, D., Grandjean, D., Sander, D., Herdener, M., Strik, W., Seifritz, E., 2008. The effect of appraisal level of processing of emotional prosody in meaningless speech. Neuroimage 42, 919–927.
Barbeau, E., Felician, O., Joubert, S., Sontheimer, A., Ceccaldi, M., 2005. Preserved visual recognition memory in an amnesic patient with hippocampal lesions. Hippocampus 15, 587–596.
Baetens, K., Steen, J., Van Overwalle, F., 2013. Involvement of the mentalizing network in social and non-social high construal. Soc. Cogn. Affect. Neurosci. 9, 817–824.
Belin, P., Zatorre, R., 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14, 2105–2109.
Beckmann, C., Jenkinson, M., Smith, S., 2003. General multilevel linear modeling for group analysis in fMRI. Neuroimage 20, 1052–1063.
Belin, P., Zatorre, R., Hoge, R., Evans, A., Pike, B., 1999. Event-related fMRI of the auditory cortex. Neuroimage 10, 417–429.
Belin, P., Zatorre, R., Ahad, P., 2002. Human temporal-lobe response to vocal sounds. Cognit. Brain Res. 13, 17–26.
Belin, P., Fecteau, S., Bedard, C., 2004. Thinking of the voice: neural correlates of voice perception. Trends Cognit. Sci. 8, 129–135.
Bestelmeyer, P., Belin, P., Grosbras, M., 2011. Right temporal TMS impairs voice detection. Curr. Biol. 21, R838–R839.
Bestelmeyer, P., Latinus, M., Bruckert, L., Rouger, J., Crabbe, F., Belin, P., 2012. Implicitly perceived vocal attractiveness modulates prefrontal cortex activity. Cerebr. Cortex 22, 1263–1270.
Bestelmeyer, P., Belin, P., Ladd, R., 2015. A neural marker for social bias toward in-group bias. Cerebr. Cortex 25, 3953–3961.
Bradlow, A., Nygaard, L., Pisoni, D., 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. Percept. Psychophys. 61, 206–219.
Bresnahan, M., Ohashi, R., Nebashi, R., Liu, W., Shearman, S., 2002. Attitudinal and affective response toward accented English. Lang. Commun. 22, 171–185.
Brown, M., Aggleton, J., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat. Rev. Neurosci. 2, 51–61.
Caffi, C., Janney, R., 1994. Toward a pragmatics of emotive communication. J. Pragmat. 22, 325–373.
Campanella, S., Belin, P., 2007. Integrating face and voice in person perception. Trends Cogn. Sci. 11, 535–543.
Cao, Y., Sebastian Contreras-Huerta, L., McFadyen, J., Cunnington, R., 2015. Racial bias in neural response to other's pain is reduced with other-race contact. Cortex 70, 68–78.
Cavallo, A., Lungu, O., Becchio, C., Ansuini, C., Rustichini, A., Fadiga, L., 2015. When gaze opens the channel for communication: integrative role of IFG and MPFC. Neuroimage 119, 63–69.
Chen, G., Chen, G., Xie, C., Li, S., 2011. Negative functional connectivity and its dependence on the shortest path length of positive network in the resting-state human brain. Brain Connect. 1, 195–206.
Chiao, J., Cheon, B., Pornpattananangkul, N., Mrazek, A., Blizinsky, K., 2013. Cultural neuroscience: progress and promise. Psychol. Inq. 24, 1–19.

Chiao, J., Mathur, V., 2010. Intergroup empathy: how does race affect empathic neural responses. Curr. Biol. 20, R478–R480.
Coupland, N., Bishop, H., 2007. Ideologised values for British accents. J. SocioLinguistics 11, 74–93.
Derwing, T.M., Munro, M.J., 1997. Accent intelligibility and comprehensibility. Stud. Sec. Lang. Acquis. 1901, 1–16.
Dupoux, E., Green, K., 1997. Perceptual adjustment to highly compressed speech: effects of talker and rate changes. J. Exp. Psychol. Hum. Percept. Perform. 23, 914–927.
Edwards, J., 1982. Language attitudes and their implications among English speakers. In: Ryan, E., Giles, H. (Eds.), Attitudes towards Language Variation. Edward Arnold, London, pp. 20–33.
Egorova, N., Shtyrov, Y., Pulvermüller, F., 2017. Brain basis of communicative actions in language. Neuroimage 125, 857–867.
Feng, W., Wu, Y., Jan, C., Yu, H., Jiang, X., Zhou, X., 2017. Effects of contextual relevance on pragmatic inference during conversation: an fMRI study. Brain Lang. 171, 52–61.
Fiske, S., Neuberg, S., 1990. A continuum of impression formation from category-based to individuating processes: influences of information and motivation on attention and interpretation. Adv. Exp. Soc. Psychol. 23, 1–74.
Freeman, J., Schiller, D., Rule, N., Ambady, N., 2010. The neural origins of superficial and individual judgments about ingroup and out-group members. Hum. Brain Mapp. 31, 150–159.
Friederici, A., 2002. Towards a neural basis of auditory sentence processing. Trends Cognit. Sci. 6, 78–84.
Friederici, A., Alter, K., 2004. Lateralization of auditory language functions: a dynamic dual pathway model. Brain Lang. 89, 267–276.
Frühholz, S., Grandjean, D., 2013. Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. Neurosci. Biobehav. Rev. 37, 24–35.
Gaab, N., Gabrieli, J.D.E., Glover, G.H., 2007. Assessing the influence of scanner background noise on auditory processing. I. An fMRI study comparing three experimental designs with varying degrees of scanner noise. Hum. Brain Mapp. 28, 703–720.
Ghazi-Saidi, L., Dash, T., Ansaldo, A., 2015. How native-like can you possibly get: fMRI evidence for processing accent. Front. Hum. Neurosci. 9, 587.
Gibson, E., Tan, C., Rutrell, R., Mahowald, K., Konieczny, L., Hemforth, B., Fedorenko, E., 2017. Don't underestimate the benefits of being misunderstood. Psychol. Sci. 28, 703–712.
Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M., Scherer, K., Vuilleumier, P., 2005. The voices of wrath: brain responses to angry prosody in meaningless speech. Nat. Neurosci. 8, 145–146.
Greenwald, A., McGhee, D., Schwartz, J., 1998. Measuring individual differences in implicit cognition: the implicit association test. J. Pers. Soc. Psychol. 74, 1464–1480.
Han, S., 2018. Neurocognitive basis of racial ingroup bias in empathy. Trends Cognit. Sci. 22 (5), 400–421.
Han, S., Ma, Y., 2016. A culture-behavior-brain loop model of human development. Trends Cognit. Sci. 19, 666–676.
Han, S., Zhang, T., Wang, H., Han, S., 2017. Neural correlates of believing. Neuroimage 156, 155–165.
Harris, L., Fiske, S., 2007. Social groups that elicit disgust are differentially processed in mPFC. Soc. Cognit. Affect Neurosci. 2, 45–51.
Hayakawa, S., Costa, A., Foucart, A., Keysar, B., 2016. Using a foreign language changes our choices. Trends Cognit. Sci. 20, 791–793.
Hehman, E., Ingbretsen, Z., Freeman, J., 2014. The neural basis of stereotypic impact on multiple social categorization. Neuroimage 101, 704–711.
Hellbernd, N., Sammler, D., 2018. Neural basis of social communicative intentions in speech. Soc. Cognit. Affect. Neurosci. 13, 604–615.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402.
Hu, C., Wang, Q., Short, L.A., Fu, G., 2012. Speech Spectrum's correlation with speakers' eysenck personality traits. PLoS One 7, e33906.
Hurt, T., Weaver, C., 1972. Negro dialect ethnocentricism and the distortion of information in the communicative process. Cent. States Speech J. 23, 118–125.
Jenkinson, M., Smith, S.M., 2001. A global optimization method for robust affine registration of brain images. Med. Image Anal. 5, 143–156.
Jiang, X., Pell, M.D., 2015. On how the brain decodes vocal cues about speaker confidence. Cortex 66, 9–34. https://doi.org/10.1016/j.cortex.2015.02.002.
Jiang, X., Pell, M.D., 2016a. Neural responses towards a speaker's feeling of unknowing. Neuropsychologia 81, 79–93.
Jiang, X., Pell, M.D., 2016b. The feeling of another's knowing: how "mixed messages" in speech are reconciled. J. Exp. Psychol. Hum. Percept. 42, 1412–1428.
Jiang, X., Pell, M.D., 2017. The sound of confidence and doubt. Speech Commun. 88, 106–126.
Jiang, X., Sanford, R., Pell, M.D., 2017. Neural systems for evaluating speaker unbelievability. Hum. Brain Mapp. 38, 3732–3749.
Jiang, X., Gossack-Keenan, K., Pell, M.D., Social Identity Alters Believability: How Speaker Accents Modulate the Brain Signatures of Processing Vocally Expressed Confidence, Under Review.
Katsumi, Y., Dolcos, S., 2017. Neural correlates of racial group bias in observing computer-animated social encounters. Front. Hum. Neurosci. 11, 632.
Kotz, S., Meyer, M., Paulmann, S., 2006. Lateralization of emotional prosody in the brain: an overview and synopsis on the impact of study design. Prog. Brain Res. 156, 286–294.
Kotz, S., Kalberlah, C., Bahlmann, J., Friederici, A., Haynes, J., 2013. Predicting vocal emotion expressions from the human brain. Hum. Brain Mapp. 34, 1971–1981.
Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 255, 606–608.
Kuhlen, A.K., Bogler, C., Swerts, M., Haynes, J.-D., 2015. Neural coding of assessing another person's knowledge based on nonverbal cues. Soc. Cognit. Affect Neurosci. 10, 729–734.
Lau, T., Cikara, M., 2017. fMRI repetition suppression during generalized social categorization. Sci. Rep. 7, 4262.
Lev-Ari, S., 2014. Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. Front. Psychol. 5, 1546.
Lev-Ari, S., Keysar, B., 2010. Why don't we believe non-native speakers? The influence of accent on believability. J. Exp. Soc. Psychol. 46, 1093–1096.
Li, T., Cardenas-Iniguez, C., Correll, J., Cloutier, J., 2016. The impact of motivation on race-based impression formation. Neuroimage 124, 1–7.
Liao, W., Chen, H., Feng, Y., Mantini, D., Gentili, C., Pan, Z., Ding, J., Duan, X., Qiu, C., Lui, S., Gong, Q., Zhang, W., 2010. Selective aberrant functional connectivity of resting state networks in social anxiety disorder. NeuroImage 52, 1549–1558.
Lippi-Green, R., 1997. English with an Accent: Language Ideology and Discrimination in the United States. Psychology Press.
Liu, P., Rigoulot, S., Pell, M., 2016. Cultural immersion alters emotion perception: neurophysiological evidence from Chinese immigrants to Canada. Soc. Neurosci. https://doi.org/10.1080/17470919.2016.1231713.
Luo, S., Li, B., Ma, Y., Zhang, W., Rao, Y., Han, S., 2015. Oxytocin receptor gene and racial group bias in empathy-related brain activity. Neuroimage 110, 22–31.
Maril, A., Simons, J., Mitchell, J., Schwartz, B., Schacter, D., 2003. Feeling-of-knowing in episodic memory: an event-related fMRI study. Neuroimage 18, 827–836.
Maril, A., Simons, J., Weaver, J., Schacter, D., 2005. Graded recall success: an event-related fMRI comparison of tip of the tongue and feeling of knowing. Neuroimage 24, 1130–1138.
Masten, C., Morelli, S., Eisenberger, N., 2011. An fMRI investigation of empathy for “social pain” and subsequent prosocial behavior. Neuroimage 55, 381–388.
Mathur, V., Harada, T., Chiao, J., 2012. Racial identification modulates default network activity for same- and other-race faces. Hum. Brain Mapp. 33, 1883–1893.
Mazziotta, J., Toga, A., Evans, A.C., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pile, B., Holmes, C., Collins, D.L., Thompson, P.M., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Goualher, G.L., Boomsma, D., Cannon, T., Kawashima, R., Mazoyer, B., 2001. A probabilistic atlas and reference system for the human brain: international Consortium for Brain Mapping ICBM. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 1293–1322.
McAleer, P., Todorov, A., Belin, P., 2014. How do you say ‘Hello’? Personality impressions from brief novel voices. PLoS One 9, e90779.
Molenberghs, P., Cunnington, R., Mattingley, J., 2012. Brain regions with mirror properties: a meta-analysis of 125 human fMRI studies. Neurosci. Biobehav. Rev. 36, 341–349.
Mulac, A., Hanley, T., Prigge, D., 1974. Effects of phonological speech foreignness upon three dimensions of attitude of selected American listeners. Q. J. Speech 60 (4), 411–420.
Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660.
Nygaard, L., Pisoni, D., 1998. Talker-specific learning in speech perception. Percept. Psychophys. 60, 355–376.
O'Reilly, J., Woolrich, M., Behrens, T., Smith, S., Johansen-Berg, H., 2012. Tools of the trade: psychophysiological interactions and functional connectivity. Soc. Cognit. Affect Neurosci. 7, 604–609.
Oleszkiewicz Pisanski, K., Lachowicz-Tabaczek, K., Sorokowska, A., 2017. Voice-based assessments of trustworthiness competence and warmth in blind and sighted adults. Psychon. Bull. Rev. 24, 856–862.
Orena, A.J., Theodore, R., Polka, L., 2015. Language exposure facilitates talker learning prior to language comprehension even in adults. Cognition 143, 36–40.
Pell, M.D., Leonard, C.L., 2003. Processing emotional tone from speech in Parkinson's disease: a role for the basal ganglia. Cognit. Affect Behav. Neurosci. 3, 275–288.
Pell, M.D., Monetta, L., Rothermich, K., Kotz, S., Cheang, H., McDonald, S., 2014. Social perception in adults with Parkinson's disease. Neuropsychology 28, 905–916.
Perrachione, T., Wong, C.P., 2007. Learning to recognize speakers of a non-native language: implications for the functional organization of human auditory cortex. Neuropsychologia 45, 1899–1910.
Perrachione, T., Chiao, J., Wong, P., 2010. Asymmetric cultural effects on perceptual expertise underlie an own-race bias for voices. Cognition 114, 42–55.
Poldrack, R., 2006. Can cognitive processes be inferred from neuroimaging data? Trends Cognit. Sci. 10, 59–63.
Poldrack, R., 2011. Inferring mental states from neuroimaging data: from reverse inference to large-scale decoding. Neuron 72, 692–697.
Rauschecker, J., Scott, S., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. 12, 718–724.
Rigoulot, S., Fish, K., Pell, M.D., 2014. Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. Brain Res. 1565, 48–62.
Rodd, J., Davis, M., Johnsrude, I., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cerebr. Cortex 15, 1261–1269.
Ryan, E., Sebastian, R., 1980. The effects of speech style and social class background on social judgments of speakers. Br. J. Clin. Psychol. 19, 229–233.
Sammler, D., Grosbras, M., Anwander, A., Bestelmeyer, P., Belin, P., 2015. Dorsal and ventral pathways for prosody. Curr. Biol. 25, 3079–3085.
Schirmer, A., 2017. Is the voice an auditory face? An ALE meta-analysis comparing vocal and facial emotion processing. Soc. Cogn. Affect. Neurosci 13, 1–13.

Schirmer, A., Kotz, S., 2006. Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cognit. Sci. 10, 24–30.
Sheng, F., Han, S., 2012. Manipulations of cognitive strategies and intergroup relationships reduce the racial bias in empathic neural responses. Neuroimage 61, 786–797.
Shkurko, A., 2013. Is social categorization based on relational ingroup/out-group opposition? A meta-analysis. Soc. Cognit. Affect Neurosci. 8, 870–877.
Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., Luca, M., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., Stefano, N., Brady, J., Matthews, P., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219.
Stanley, D., Soko-Hessner, P., Fareri, D., Perino, M., Delgado, M., Banaji, M., Phelps, E., 2012. Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 744–753.
Sumner, M., 2015. The social weight of spoken words. Trends Cognit. Sci. 19, 238–239.
Tesink, C., Petersson, K., van Berkum, J., van den Brink, D., Buitelaar, J., Hagoort, P., 2009. Unification of speaker and meaning in language comprehension: an fMRI study. J. Cognit. Neurosci. 21, 2085–2099.
Tsantani, M., Belin, P., Paterson, H., McAleer, P., 2016. Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. Perception 45, 946–963.
Van Berkum, J.J., Van den Brink, D., Tesink, C., Kos, M., Hagoort, P., 2008. The neural integration of addresser and message. J. Cognit. Neurosci. 20, 580–591.
Van den Brink, D., Van Berkum, J.J., Bastiaansen, M., Tesink, C., Kos, M., Buitelaar, J., et al., 2012. Empathy matters: ERP evidence for inter-individual differences in social language processing. Soc. Cognit. Affect Neurosci. 7, 173–183.
Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30, 829–858.
Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48, 564–584.
Van Overwalle, F., Baetens, K., Mariën, P., Vandekerchkhove, M., 2014. Social cognitive and the cerebellum: a meta-analysis of over 350 fMRI studies. Neuroimage 86, 554–572.
Vandenberghe, R., Nobre, A., Price, C., 2006. The response of left temporal cortex to sentences. J. Cognit. Neurosci. 14, 550–560.
Von Kriegstein, K., Giraud, A., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. Neuroimage 22, 948–955.
Wildgruber, D., Ackermann, H., Kreifelts, B., Ethofer, T., 2006. Cerebral processing of linguistic and emotional prosody: fMRI studies. Prog. Brain Res. 156, 249–268.
Woolrich, M., Ripley, B., Brady, M., Smith, S., 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. Neuroimage 14, 1370–1386.
Woolrich, M., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S., 2004. Multilevel linear modelling for fMRI group analysis using Bayesian inference. Neuroimage 21, 1732–1747.
Worsley, K., Friston, K., 1995. Analysis of fMRI time-series revisted-Again. Neuroimage 2, 173–181.
Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. Neuroimage 15, 1–15.
Wu, D., Loke, I.C., Xu, F., Lee, K., 2011. Neural correlates of evaluations of lying and truth-telling in different social contexts. Brain Res. 1389, 115–124.
Xu, X., Zuo, X., Wang, X., Han, S., 2009. Do you feel my pain? Racial group membership modulates empathic neural responses. J. Neurosci. 29, 8525–8529.
Zuo, X., Han, S., 2013. Cultural experiences reduce racial bias in neural responses to others' suffering. Cult. Brain 1, 34–46.