Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies

Michel Belyk¹, Steven Brown¹

¹Department of Psychology, Neuroscience & Behaviour, McMaster University
Hamilton, Ontario, Canada

RUNNING TITLE: Prosody meta-analysis

Correspondence to:
Michel Belyk
Department of Psychology, Neuroscience & Behaviour
McMaster University
1280 Main St. West
Hamilton, ON, L8S 4M9
Canada
Phone: 1 (905) 525 9140 extension 23892
e-mail: belykm@mcmaster.ca

© The Author (2013). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com
Prosody refers to the melodic and rhythmic aspects of speech. Two forms of prosody are typically distinguished: “affective prosody” refers to the expression of emotion in speech, whereas “linguistic prosody” relates to the intonation of sentences, including the specification of focus within sentences and stress within polysyllabic words. While these two processes are united by their use of vocal pitch modulation, they are functionally distinct. In order to examine the localization and lateralization of speech prosody in the brain, we performed two voxel-based meta-analyses of neuroimaging studies of the perception of affective and linguistic prosody. There was substantial sharing of brain activations between analyses, particularly in right-hemisphere auditory areas. However, a major point of divergence was observed in the inferior frontal gyrus: affective prosody was more likely to activate Brodmann area 47, while linguistic prosody was more likely to activate the ventral part of area 44.

Keywords: affective prosody, linguistic prosody, speech, emotion, ALE meta-analysis, brain imaging, inferior frontal gyrus
1. Introduction

Prosody comes from the Greek *prosodia*, meaning “sung to music” (Pearsall, Hanks, Soanes & Stevenson, 2005). Speech prosody therefore refers to the song-like vocal modulations that accompany speech. For this reason, it is often considered to be “the music of speech” (Wennerstrom, 2001). The pitch modulations associated with speech prosody convey two broad categories of information. On the one hand, pitch modulations convey information about a speaker’s emotional state (Fairbanks & Pronovost, 1938), what has been referred to as “emotional” or “affective” prosody (Monrad-Krohn, 1947). On the other hand, they provide cues regarding syntax and pragmatics (Beach, 1991), what has been referred to as “intrinsic” or “linguistic” prosody (Monrad-Krohn, 1947). While these two types of prosody are functionally distinct, they rely on a common set of acoustic cues related to pitch, loudness, tempo, and voice quality (Fonagy, 1978; Juslin & Laukka, 2003). The sharing of acoustic parameters by these two processes suggests that they might rely on a common system for the perception of pitch but that this pitch information may be fed into distinct systems for processing either emotion (affective prosody) or syntax/pragmatics (linguistic prosody).

Affective prosody conveys a speaker’s emotional state largely through global changes in pitch height and loudness, although other acoustic features also serve to disambiguate emotional states (Banse and Sherer, 1996). Emotional expressions can take the form of “affect bursts” (Schröder, 2003) that have emotional but not semantic meaning (e.g., “Yuck!”) or can occur concurrently with normal speech. Affective prosody conveys a broad range of emotional states (Sauter and Scott, 2007) that can be
recognized across cultures without prior experience (Sauter et al., 2010; Scherer et al., 2001), much like facial expressions (Ekman et al., 1969).

Linguistic prosody uses local increases in pitch height and/or loudness to signal features like word stress (e.g., CONtent vs. conTENT; Gay, 1978), sentence focus (e.g., two WHITE shirts vs. TWO white shirts; Ladd and Morton, 1997), segmentation of the speech stream into phrases (Jusczyk et al., 1992), broad pragmatic categories of utterances (modality), such as declarative vs. interrogative sentences (Xu and Xu, 2005), and the standard intonational melodies that are used as part of mother-infant communication (Fernald, 1992) as well as communication between adults. The conventions of linguistic prosody vary across languages and are important contributors to the melody and rhythm of speech. It is for this reason that deviations from standard prosody contribute to the impression of a foreign accent (de Mareüil and Vieru-Dimulescu, 2006).

Early investigations into the neural basis of speech prosody analyzed neurological cases of patients suffering from strokes. These studies focused overwhelmingly on the lateralization of prosody, especially compared to the well-accepted left-hemisphere dominance for the lexicosyntactic aspect of language. A major finding of these early studies was that the perception of affective prosody was impaired in patients with unilateral right-hemisphere lesions (Gorelick and Ross, 1987; Ross, 1981). However, these studies did not examine patients with left-hemisphere lesions, and studies that have since done so have reported deficits in patients with both types of unilateral lesions (Pell, 1998; Trauner et al., 1996). Similarly, deficits in the perception of linguistic prosody have been reported in patients with lesions in both the left (Pell and Baum, 1997) and
right (Weintraub et al., 1981) hemispheres. A meta-analysis of this literature revealed that both affective and linguistic prosody are impaired by damage to either hemisphere, although damage to the right hemisphere tends to have a larger impact on affective prosody and the left hemisphere on linguistic prosody (Witteman, Van Ijzendoorn, Van de Velde, Van Heuven, & Schiller, 2011).

Neurological studies have generally been conducted with patients having a diverse set of lesions and have seldom reported the location of lesions beyond the level of the hemisphere or lobe. Therefore, the neurological literature does not permit an examination of localization hypotheses at a finer scale than the lobe. Interestingly, transcranial magnetic stimulation of healthy individuals can induce deficits in the perception of affective prosody when applied to either the left or right inferior frontal gyrus (IFG; Hoekert et al., 2010).

Neuroimaging studies have been similarly inconclusive with respect to the hemispheric lateralization of prosody perception. The literature has variably reported unilateral or bilateral activations for affective prosody (Bach et al., 2008; Ethofer et al., 2009; Wildgruber et al., 2005) and linguistic prosody (Meyer et al., 2002; Strelnikov et al., 2006). Despite these inconsistencies in lateralization, neuroimaging studies have contributed to the broader localizationist account of prosody perception.

Neural models of affective-prosody perception (Ethofer et al., 2006; Schirmer and Kotz, 2006) suggest that low-level acoustic analyses are performed in the posterior superior temporal gyrus (STG) – in what has been called the “emotional voice area” (Ethofer et al., 2012) – and the superior temporal sulcus (STS). Similarly, more recent models suggest that acoustic processing is performed in the middle part of the superior
temporal sulcus (mSTS; Belin et al., 2000), that identification of vocally expressed emotions is performed in either the anterior (Kotz & Paulmann, 2011) or posterior (Brück, Kreifelts, & Wildgruber, 2011) STG/STS, and that explicit evaluation of vocally-expressed emotions is performed by inferior frontal regions (Wildgruber et al., 2009). Passive perception of prosody reliably activates the STG (Dietrich et al., 2008; Humphries et al., 2005). Posterior temporal areas are proposed to project to inferior frontal regions for explicit evaluation of emotional meaning when such evaluation is task-relevant. While studies of both affective and linguistic prosody routinely report activations in Broca’s area (Gandour et al., 2003a; Gandour et al., 2003b), Schirmer and Kotz (2006) proposed that a region anteroventral to Broca’s area – the IFG pars orbitalis (Brodmann area [BA] 47) – may be specifically involved in the perception of affective prosody. A meta-analysis of the imaging literature on the perception of affective prosody supports the involvement of the IFG pars orbitalis when attention is directed towards affective prosody rather than away from it and the IFG pars triangularis (BA 45) whether or not attention is directed towards affective prosody (Witteman, Van Heuven, and Schiller, 2012).

The perception of prosody stimulates additional regions beyond the superior temporal and inferior frontal gyri (Brück et al., 2011; Buchanan et al., 2000). Studies of affective and linguistic prosody routinely report activations in speech-related areas – even when contrasted with other speech-perception tasks – including the anterior cingulate cortex (ACC; Doherty et al., 2004; Frühholz et al., 2011), inferior parietal lobule (IPL; Gandour et al., 2003a; Johnstone et al., 2006), anterior insula (Ethofer et al., 2009; Meyer et al., 2002), and basal ganglia (Bach et al., 2008; Meyer et al., 2004).
Given the inconsistencies in both the neurological and neuroimaging literatures, we sought to clarify the localization of prosody perception in the brain by performing a statistical meta-analysis of published neuroimaging studies of affective and linguistic prosody either separately, in contrast, or in conjunction using the “activation likelihood estimation” (ALE) method (Eickhoff et al., 2011; Turkeltaub et al., 2002). The goal was to assess whether these two functions are mediated by shared or distinct brain networks. The major predictions were that these functions should show commonalities in posterior temporal areas that process the acoustic features of vocal pitch, but that differences should be seen in higher-level areas in the frontal lobe that generate distinct interpretations of these pitch modulations.

2. Methods

2.1 Inclusion criteria

A meta-analysis of published neuroimaging studies of affective and linguistic prosody was performed using ALE meta-analysis (Turkeltaub et al., 2002) in order to compare areas of brain activation across these functions. Published articles were retrieved in February 2012 by searches in the Web of Knowledge database using the search terms “prosody + fMRI” and “prosody + PET”. The reference sections of resultant studies were searched for additional studies. Experiments in which subjects made emotional judgments were classified as “affective prosody”, while studies in which subjects made judgments based on word stress, focus, syntax, or modality were classified as “linguistic prosody”. 
Our inclusion criteria for the studies were: 1) that brain scanning was performed using either functional magnetic resonance imaging (fMRI) or positron emission tomography (PET); 2) that papers reported activation foci in the form of standardized stereotaxic coordinates in either Talairach or Montreal Neurological Institute (MNI) space; 3) that subjects were healthy adults (thereby excluding results from clinical populations); 4) that subjects made active judgments about the affective or linguistic prosody of auditorily-presented speech stimuli; 5) that the analyses included a high-level contrast against a suitable control condition so as to remove the influence of low-level phonological processing (e.g., passive listening or gender discrimination); and 6) that results from the entire scanned volume were reported (thereby excluding studies reporting region-of-interest analyses only). Due to the large number of studies with only partial brain coverage, we performed a separate analysis with the additional criterion 7) that the entire brain-volume was imaged (thereby excluding studies with an insufficient field of view to encompass the whole brain). This criterion is discussed further in section 2.2.

Our searches yielded 29 independent experiments conducted in German, English, French, Mandarin, Japanese, and Russian (see Supplementary Tables 1 and 2 for details). Wherever studies reported multiple experiments from the same group of subjects, the contrasts were included together as a single study. Similarly, for studies that reported the results of more than one subject-group, each group was treated separately, in accordance with the approach of Turkeltaub et al. (2011). Separate analyses were conducted for affective prosody (n=19 experiments) and linguistic prosody (n=10). GingerALE 2.1 was used for all analyses and to convert MNI coordinates to Talairach coordinates. The ALE...
results were registered onto a Talairach-normalized template brain using Mango
(ric.uthscsa.edu/mango). All analyses were corrected for multiple comparisons using the
False Discovery Rate $p < 0.05$ and cluster threshold $k = 10$.

2.2 Brain coverage

ALE meta-analysis is an empirical technique for the analysis of brain imaging
studies (Turkeltaub et al., 2002). Each focus of activation is modeled as a three-
dimensional Gaussian probability distribution whose width is determined by the size of
the subject-group so as to reflect increasing uncertainty with decreasing sample size
(Eickhoff et al., 2009). Maps of activation likelihoods are created for each study by
taking the maximum probability of activation at each voxel. A random-effects analysis
tests for the convergence of activations across studies against a null hypothesis of
spatially independent brain activations.

Due to the limited brain coverage of many of the studies included in our dataset,
we modified the standard ALE method in order to test the null hypothesis of spatially
independent brain activations within the brain volume that was imaged in all of the
included studies. Standard ALE analyses mask the brain volume to grey matter.
Activation foci are unlikely to originate from ventricles or white matter. Therefore, in
order to avoid skewing the empirical null distribution – and overestimating any effects –
this portion of brain space must be excluded (Eickhoff et al., 2009). Similarly, activation
foci cannot originate from outside the field of view for a given study, and so this region
must therefore be excluded from the analysis. We therefore further restricted the analyses
to the portion of the brain-volume that was imaged in all studies meeting our inclusion
criteria. This area extended from $z = -6$ to $z = 38$ in Talairach space (see horizontal red
lines in Figure 1). A second set of whole-brain analyses was performed to assess convergence beyond the restricted volume of coverage. Whole-brain analyses only included studies that imaged the entire brain volume. This additional inclusion criterion reduced the number of studies to 10 and 4, respectively, for affective and linguistic prosody. For all figures and tables, all 29 experiments contributed to analyses within the restricted range. Only those experiments with full brain coverage contributed to analyses outside this range.

2.3 Conjunctions and contrasts

In addition to separate analyses, we performed a statistical conjunction (Nichols et al., 2005) of the meta-analyses in order to determine which areas, if any, were common to affective and linguistic prosody. Direct contrasts were performed to determine which areas were specific to each of these two functions. Because there were many more studies of affective prosody than linguistic prosody in the dataset – which may bias the results – we also report the number and percentage of studies of affective prosody and linguistic prosody that contribute to each of the ALE foci. Due to the small number of studies covering the whole brain, direct contrasts are reported for the restricted analysis only.

2.4 Post hoc analysis of working memory demands based on task-type

The studies included in the meta-analyses used tasks that fall into two broad classes: identification tasks and same/different tasks. Subjects performing an identification task are presented with an auditory stimulus and are required to identify – from a limited set of possible responses – which emotion or intonation is being presented. Subjects performing a same/different task are presented with pairs of stimuli and are
required to indicate whether the same emotion or intonation occurs in both presentations. To the extent that the latter task requires subjects to maintain a representation of the first stimulus-presentation long enough to perform a comparison with the second, it may impose greater demands on working memory than an identification task. Among the studies included in our meta-analyses, affective prosody experiments were much more likely to use identification tasks or similar tasks with a low working memory load (16 out of 19), while studies of linguistic prosody were more evenly divided (4 and 6 low and high working memory load, respectively). We therefore compared experiments of linguistic prosody containing putatively low vs. high working memory load as, estimated from task demands, in order to account for areas of convergence that may be more reflective of working memory demands than prosody perception per se.

3. Results

We performed individual ALE analyses of affective and linguistic prosody. Due to the preponderance of studies with functional coverage limited to the perisylvian region alone, we performed two parallel analyses for each function, one restricted to the volume covered by all studies in the dataset (in order to avoid violating the assumptions of the ALE method) and a second, whole-brain analysis exclusively for those studies that reported whole-brain coverage. Results from both the restricted and whole brain analyses are combined in all figures and tables. Figure 1 presents the location of the major ALE foci for each analysis, and Table 1 provides the Talairach coordinates and cluster sizes for each ALE focus. Results will first be presented for analyses of each function.
separately, followed by a conjunction of analyses to identify shared regions, and finally
direct contrasts to identify regions specific to each function.

***Insert Figure 1 and Table 1 here***

Affective prosody activated both audio-vocal and limbic areas. Audio-vocal activations were observed in right pSTG, bilateral aSTG, supplementary motor area (SMA), IFG pars opercularis (BA 44), pars triangularis (BA 45) and supramarginal gyrus, right middle temporal gyrus (MTG), cerebellum, and middle frontal gyrus (BA 9, BA 10), left caudate nucleus and thalamus. Presumed emotion-related activations were observed in limbic areas, including bilateral IFG pars orbitalis (BA 47), left amygdala, ventral anterior insula and ventral putamen, right parahippocampal gyrus (BA 28), and subcallosal gyrus (BA 34). Importantly, the frontal language areas that are most widely discussed in this literature showed bilateral activity.

In contrast to this limbic profile for affective prosody, linguistic prosody showed ALE foci primarily in speech and language areas. These included bilateral IFG pars opercularis (BA 44), pSTG, supramarginal gyrus (BA 40), middle frontal gyrus, right SMA, IFG pars orbitalis, primary auditory cortex (BA 41) and the left caudate nucleus. Non-language-related foci were observed in the bilateral insula and cerebellum as well as in the right claustrum and primary visual cortex. As with affective prosody, the ALE foci in frontal perisylvian language areas were present bilaterally.

***Insert Table 2 here***
Next, we compared the functions using conjunctions so as to identify areas of overlap versus areas of function-specificity (see Figure 1 and Table 2). Conjunction analyses demonstrated that affective prosody shared common areas with linguistic prosody. As predicted, affective and linguistic prosody showed overlapping activations in the right superior temporal gyrus (BA 22). Other areas of overlap included the bilateral supramarginal gyrus, right middle frontal gyrus, left insula, and midline SMA.

***Insert Figure 2 and Table 3 here***

In order to identify regions that were specific to each condition, we performed direct contrasts (see the right panel of Figure 2 and Table 3). Affective prosody had a stronger association with activation in the left IFG pars orbitalis (BA 47) and thalamus as well as right pSTG (BA 22) and MTG (BA 21). Linguistic prosody had a stronger association with activation in the left pSTG, bilateral middle frontal gyrus (BA 9), bilateral IFG pars opercularis (BA 44), right supramarginal gyrus (BA 40), claustrum, and midline primary visual cortex.

***Insert Table 4 here***
As an additional analysis, we divided the studies of linguistic-prosody perception into those with putatively high versus low verbal working-memory load, as estimated by task demands (Table 4). Higher working memory load was associated with increased activation in the bilateral middle frontal gyrus (part of the dorsolateral prefrontal cortex) and right STG. Given that the literatures under review here were not orthogonal with respect to working memory demands, differences between individual ALE analyses in the right STG and middle frontal gyri should be interpreted with caution.

4. Discussion

The primary objective of this study was to use meta-analytic techniques to help clarify which brain regions are reported consistently in studies of affective and linguistic prosody perception in light of inconsistency and confusion in both the neurological and neuroimaging literatures. We examined the functional neuroimaging literatures related to affective and linguistic prosody individually and then jointly using conjunction and contrast methods. The results revealed both shared and distinct components of the networks involved in these processes, reflecting both the perception of vocal-pitch modulation and its functional interpretation.

Our strongest prediction of overlap between the two functions was for auditory association areas in the pSTG. Interestingly, the right pSTG has been dubbed the “emotional voice area” by researchers of affective prosody (Ethofer et al., 2012). In confirmation of this area’s role in emotional voice perception, we observed convergence centered in right Heschl’s gyrus and extending into the pSTG for studies of affective
prosody. However, we observed a similar area of convergence bilaterally for linguistic prosody as well as for verbal working memory, and the pSTG is commonly reported in studies of music perception as well (Brown et al., 2004; Zatorre et al., 1994). Indeed, Wiethoff et al. (2008) observed that activation in this region could be explained entirely by the acoustic parameters of the stimuli. The pSTG appears to respond to a variety of types of auditory stimuli and may not be specific to emotional voices. Emotional voices may simply contain a larger degree of pitch modulation than the neutral voices that are typically used as baseline stimuli in many studies of affective prosody.

In addition to demonstrating overlap in right auditory areas, the conjunction analysis revealed convergence across functions in the SMA, a motor structure involved in speech production. Electrical stimulation of the anterior portion of the left (but not right) SMA elicits vocalization (Fried et al., 1991). Lesions to this area can cause aphasic symptoms (Fontaine et al., 2002) and akinetic mutism (Nagaratnam et al., 2004). The individual ALE meta-analyses of affective prosody and linguistic prosody suggested a role of the right SMA in those functions as well. Surprisingly, convergence between studies was not observed in the ACC just ventral to the SMA. Animal models of vocal control demonstrate an important role of the ACC in top-down control of the vocal-motor nuclei in the brain stem (Jürgens, 2002), and ALE meta-analysis of human neuroimaging studies of vocalization show foci in this region for both spoken and sung utterances (Brown et al., 2009). Indeed, several of the studies included in these meta-analyses reported activations in the ACC (Bach et al., 2008; Doherty et al., 2004; Gandour et al., 1998), and yet ALE foci in this region did not reach significance for any analysis. One likely explanation is that much of the ACC lies outside the volume covered in our
“restricted” analyses and that our whole-brain analyses had too little power to detect convergence in this area. In addition, the anatomy of the ACC is variable across individuals (Paus et al., 1996), and it is therefore possible that differences between subjects in cingulate anatomy resulted in subtle variability in the localization of foci between studies.

Unlike the result in auditory areas, substantial divergence was observed in inferior frontal regions. More specifically, affective prosody activated the IFG pars orbitalis (BA 47) bilaterally while linguistic prosody activated the IFG pars opercularis (ventral BA 44) bilaterally. The absence of overlap in the inferior frontal region suggests that activations here do not simply relate to some aspect of pitch processing alone, but may instead reflect the different classes of information that listeners extract from affective versus linguistic cues in speech prosody. Notably, Wildgruber et al., (2004) compared affective and linguistic prosody perception directly in an fMRI study. These authors observed a similar localization for both functions in inferior frontal regions.

We observed a small number of areas that were uniquely associated with each function of interest. The IFG pars orbitalis (BA 47), which was associated most strongly with affective prosody, is distinct from adjacent Broca’s area in both cytoarchitecture (Brodmann, 1909/1994) and structural connectivity. The homologous region in macaques, area 47/12, receives projections from both limbic regions and the homologue of Broca’s area (Petrides and Pandya, 2001). Area 47/12 is part of an orbitofrontal network that receives input from sensory areas, including auditory, visual, somatosensory, olfactory, visceral and gustatory cortices, as well as limbic areas such as the amygdala, subiculum, entorhinal cortex, and perirhinal cortex. This same network
projects to the hypothalamus and periaqueductal grey by way of the ventromedial prefrontal cortex (Price, 1999). Diffusion tensor imaging in humans reveals a similar pattern. The IFG pars orbitalis is connected to auditory and visual areas via the inferior occipitofrontal fasciculus and middle longitudinal fasciculus (Turken and Dronkers, 2011). The frontal operculum adjacent to BA 47 is connected to the amygdala and septal region (Anwander et al., 2007) and plays a role in emotion regulation in conjunction with the amygdalae and nucleus accumbens (Wager et al., 2008). Patients with lesions in this region and the adjacent orbitofrontal cortex have deficits in recognizing emotions in others as well as changes in behavior and subjective emotional experience (Hornak, Rolls and Wade, 1996). This region is consistently active when subjects experience particular emotions or when they perceive emotions in either the auditory or visual domain (Lindquist et al., 2012). The IFG pars orbitalis may therefore be well situated to act as an interface between limbic and sensorimotor networks, as would be necessary for affective prosody perception. Indeed, given the diverse sensory information available to this region, it is not surprising that it is involved in the perception of emotional faces and gestures as well (Lotze et al., 2006; Sprengelmeyer et al., 1998).

Linguistic prosody, by contrast, was associated most strongly with the IFG pars opercularis (BA 44). Ventral BA 44 is associated with lexicosyntactic function. Functional MRI studies have shown that syntactic processing activates ventral IFG pars opercularis (Friederici et al., 2000; Heim et al., 2003a). In contrast, other linguistic functions, such as phonological processing, activate a locus in dorsal BA 44 (Heim et al., 2003b; Papoutsi et al., 2009). Given that linguistic prosody plays a role in syntactic
disambiguation (Beach, 1991), it is perhaps not surprising that this suprasegmental
element of speech shares brain areas with syntactic processing.

4.2 *Lateralization versus localization*

Both the neurological and neuroimaging literatures on the perception of prosody
are concerned primarily with the lateralization of function in temporal and frontal
language areas. In agreement with this literature, we observed consistent right-
hemisphere lateralization in temporal-lobe auditory areas. Importantly, we observed this
pattern of lateralization for both affective and linguistic prosodies, constituting a region
of overlap between these functions. In contrast, our results did not support a consistent
lateralization in the frontal lobe for either affective or linguistic prosody. Our meta-
analyses instead demonstrated that bilateral inferior frontal activations were likely to be
reported by neuroimaging studies of both functions, although in non-overlapping regions.
While direct contrasts between conditions appeared to support the lateralization of
affective prosody to the left inferior frontal gyrus, our primary analyses demonstrated that
affective prosody perception did in fact activate right inferior frontal regions as well.
Both affective and linguistic prosody activated bilateral (although distinct) inferior frontal
regions, as demonstrated by Figure 1. However, this does not preclude the interpretation
that some functional aspect of the task may influence patterns of lateralization. It has
been proposed that one contributor to the frequent, but inconsistent, lateralization of
speech prosody, especially in temporal-lobe auditory areas, is that the window of
temporal integration of pitch information differs between the two hemispheres (Buchanan
et al., 2000) such that the left hemisphere processes relatively fast frequency modulations
and the right hemisphere relatively slow modulations (Zatorre, 2001).
4.1 Task-type

The middle frontal gyrus (part of the dorsolateral prefrontal cortex) and STG were associated with task-related differences in working-memory load in the linguistic-prosody meta-analysis. Activations in this region were more prominent in studies of linguistic prosody than affective prosody perception. This may be due to a greater proclivity towards experiments with high verbal working-memory demands in that literature. Studies of linguistic prosody used methods with either a high working memory load, namely same/different tasks, or with a low working memory load, namely forced-choice identification tasks. In comparison, studies of affective prosody used primarily tasks with low verbal working-memory demands. This methodological difference might account for the increased likelihood of observing activation in the middle frontal gyrus for linguistic prosody compared with affective prosody. This finding is corroborated by a meta-analysis that explicitly examined verbal working-memory demands (Chein et al., 2002). Note that this analysis was conducted to detect confounds in our primary analyses and should not be taken as an analysis of working memory per-se.

4.3 Production and perception

To the best of our knowledge, only one study has compared functional activations between perception and production of prosody, and it did so for both linguistic and affective prosody (Aziz-Zadeh et al., 2010). While that study did not observe activation in the IFG pars orbitalis that we described for affective prosody, it did observe activation in the left IFG pars opercularis for the production and perception of both affective and linguistic prosodies. Our meta-analyses revealed ALE foci in this region, although the localization varied for each function. The IFG pars opercularis may be an important point
of interaction for affective and linguistic prosody. More specifically, the IFG pars opecularis is purported to be a “mirror neuron” area involved in both the production and perception of actions (Aziz-Zadeh et al., 2006). This area that may be structurally connected with the primary motor cortex (Greenlee et al., 2004; Simonyan et al., 2009). It may therefore constitute an area of convergence for affective and linguistic prosody en route to the motor cortex.

4.4 Prosody networks

A number of models have proposed temporo-frontal networks for prosody perception based on the activation patterns for affective prosody (Ethofer et al., 2006; Schirmer & Kotz, 2006). Two recent models have suggested that prosody perception occurs in three stages: 1) acoustic analysis in the voice-selective areas of the mSTS (Belin et al., 2000), 2) identification of vocally expressed emotion in the aSTG (Kotz & Paulmann, 2011) or pSTG (Brück et al., 2011) and 3) explicit evaluation of prosody in the IFG. These models agree with one another in most respects, with the exception of the localization of temporal regions specific for affective voices. While another prosody meta-analysis (Witteman et al., 2012) supported the localization of Brück et al. (2012) to the pSTG, our results supported the role of both aSTG and pSTG in affective prosody processing. Notably, we observed ALE foci in bilateral aSTG for affective prosody only, not for linguistic prosody. However these foci did not survive a direct contrast between the two functions. Due to the lower power of the linguistic-prosody analysis relative to affective prosody, it cannot be concluded from the data that either the aSTG or pSTG is specific to affective prosody.
Conjunction analysis revealed several areas of common activation between affective and linguistic prosody. Among these were the right auditory association cortex, which is specialized for the fine-grained analysis of pitch (Zatorre & Gandour, 2008), left anterior insula, which is anatomically connected to the entire extent of the IFG (spanning the pars opercularis, pars triangularis and pars orbitalis; Catani et al., 2012), and the somatotopic oro-laryngeal portion of the SMA (Fried et al., 1991). This group of regions is likely involved in audio-vocal functioning generally, rather than prosody specifically.

Affective and linguistic prosody do not generally occur in isolation but rather in parallel with speech. A focus of future research should be to further develop network models of prosody perception and to extend these models to incorporate production with the aim of integrating these networks with extant models of speech. For example, the “Directions into Velocities and Articulators” (DIVA) model (Golfinopoulos et al., 2010) is a well established model of speech production that describes how intended speech sounds are converted into articulatory movements that ultimately result in the production of speech. Such a set of mechanisms should, in theory, accommodate the production of the pitch-based cues that are used for affective and linguistic prosodies.

The IFG pars opercularis locus observed for linguistic prosody is part of Broca’s area (and Broca’s homologue) and is therefore already a component of most neural models of speech. However, the expression of emotion is acoustically similar whether it occurs without language in the form of affect bursts such as laughter and crying (Schröder, 2003) or with language in the form of affective prosody (Banse and Sherer, 1996). Affective prosody may therefore require the integration of an evolutionarily ancestral subcortical system for affective communication found in monkeys (Jürgens,
2009) with the evolutionarily recent cortical system for speech and language that is found only in humans. We suggest that the IFG pars orbitalis (BA 47) may function as such an interface between emotion and vocalization, although others have proposed that the ACC serves this function as well (Jürgens, 2009). One caveat to this proposal is the suggested role of the IFG pars orbitalis in other functions. This region has previously been reported in neuroimaging studies of both linguistic (Fiez, 1997) and musical semantics (Levitin and Menon 2003) as well as in pitch memory (Zatorre et al., 1994). Price (1999) noted that the orbital region of the macaque, including BA 47/12, is cytoarchitectonically diverse. Further research is needed to search for potential functional subdivisions within this region.

5. Limitations

A potential limitation of our analysis is that our dataset included more studies of affective prosody than linguistic prosody. This unbalanced design may have introduced some bias into the data and limited the inferences that could be made from it. We attempted to mitigate this limitation by checking the number of studies that contribute to each of the foci in our contrasts.

Our analysis of working memory load relied on a small and unbalanced sample of studies of linguistic prosody perception. Furthermore, our division into high and low working memory load was confounded with the distinction between task-driven effects and stimulus-driven effects discussed by Witteman et al. (2012). For these reasons, we stress that our working memory results are provisional and are intended only to aid in the interpretation of the other analyses.
6. Conclusion

We meta-analyzed the literatures on the neural correlates of two pitch-based paralinguistic functions. The results provide mixed support for hemispheric lateralization of speech prosody, with greater lateralization seen in temporal-lobe auditory areas than in frontal-lobe evaluative areas. Instead, the results support a localizationist account based on differentiation of the two prosodic functions in the inferior frontal gyrus. Linguistic prosody is associated with a portion of the IFG pars opercularis that is involved in syntactic processing. Affective prosody is associated with the IFG pars orbitalis, which is connected with both limbic and speech-motor areas, making it a good candidate as an interface between emotion and voice.

6. Funding

This work was funded by a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to S.B.
References


study. *Neuroreport, 19*(18), 1751-5.


components of language in the right hemisphere. *Archives of Neurology, 38*(9), 561-69.


Legends

**Figure 1**: Sagittal sections showing major foci for the individual ALE meta-analyses for affective prosody (red) and linguistic prosody (green) as well as the statistical conjunction of the two (blue). These slices demonstrate the bilateral involvement of inferior frontal regions for affective and linguistic prosody perception. The figure also demonstrates the clear segregation of functions within the inferior frontal gyrus as well as sharing in the right auditory cortex. Red lines demarcate the limits of the “restricted” analysis (z = -6 to z = 38): foci within the red lines were generated by the restricted analysis (which included all studies), while foci outside the red lines were generated by the whole-brain analysis (including only those studies that reported whole-brain coverage).

**Figure 2**: The left panel shows the ALE foci for affective prosody (AP, red) and linguistic prosody (LP, green) registered onto axial sections. The right panel shows two direct contrasts, and highlights areas unique to each function. Affective prosody is uniquely associated with the IFG pars orbitalis (BA 47), while linguistic prosody is uniquely associated with the ventral IFG pars opercularis, (BA 44).

**Table 1**: Talairach coordinates of likelihood maxima and cluster sizes for individual ALE analyses of affective prosody and linguistic prosody perception, respectively. Results from both the restricted and whole-brain analysis are combined. IFG: inferior frontal gyrus; MOG: middle occipital gyrus; MTG: middle temporal gyrus; PHG: parahippocampal gyrus; SMA: supplementary motor area; SMG: supramarginal gyrus; STG: superior temporal gyrus.
Table 2: Statistical conjunction demonstrates areas of commonality between affective prosody and linguistic prosody.

Table 3: Pairwise contrasts demonstrate areas of activation unique to affective prosody and linguistic prosody. The number of studies of affective and linguistic prosody contributing to each locus corroborates the ALE results.

Table 4: Meta-analysis of verbal working memory. Studies of linguistic prosody were divided into those with high vs. low working-memory demands based on task type. The bilateral middle frontal gyrus and right STG are more likely to be reported in studies with high verbal working-memory demands. Contrasts between individual ALE meta-analyses must be interpreted cautiously to avoid falsely attributing foci in these areas to prosody perception. The number of studies with low and high working memory loads contributing to each locus corroborates the ALE results.

Supplementary Table 1: Details of studies included in the two meta-analyses. Listed for each study are the analyses to which each study contributed, the number of subjects, type of task and control conditions, verbal working-memory demands, and brain coverage.

Supplementary Table 2: List of references for the studies included in the two meta-analyses.

Supplementary Table 3: A small and unbalanced number of the studies included in our analyses reported PET data. We reanalyzed the data excluding PET studies. The results of this supplementary analysis agree with the larger analysis that included PET studies.
Figure 1: Sagittal sections showing major foci for the individual ALE meta-analyses for affective prosody (red) and linguistic prosody (green) as well as the statistical conjunction of the two (blue). These slices demonstrate the bilateral involvement of inferior frontal regions for affective and linguistic prosody perception. The figure also demonstrates the clear segregation of functions within the inferior frontal gyrus as well as sharing in the right auditory cortex. Red lines demarcate the limits of the “restricted” analysis ($z = -6$ to $z = 38$): foci within the red lines were generated by the restricted analysis (which included all studies), while foci outside the red lines were generated by the whole-brain analysis (including only those studies that reported whole-brain coverage).
Figure 2: The left panel shows the ALE foci for affective prosody (AP, red) and linguistic prosody (LP, green) registered onto axial sections. The right panel shows two direct contrasts, and highlights areas unique to each function. Affective prosody is uniquely associated with the IFG pars orbitalis (BA 47), while linguistic prosody is uniquely associated with the ventral IFG pars opercularis, (BA 44).
Table 1

<table>
<thead>
<tr>
<th>Brain Region</th>
<th>Affective Prosody</th>
<th>Linguistic Prosody</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td><strong>Right Hemisphere</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Frontal Lobe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG pars triangularis (BA 45)</td>
<td>46</td>
<td>22</td>
</tr>
<tr>
<td>Insula (BA 13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Insula</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG pars orbitalis (BA 47)</td>
<td>48</td>
<td>14</td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td>48</td>
<td>16</td>
</tr>
<tr>
<td>SMA (BA 6)</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 10)</td>
<td>34</td>
<td>36</td>
</tr>
<tr>
<td>IFG pars opercularis (BA 44)</td>
<td>54</td>
<td>8</td>
</tr>
<tr>
<td><strong>Temporal Lobe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>aSTG (BA 22)</td>
<td>54</td>
<td>0</td>
</tr>
<tr>
<td>pSTG (BA 22)</td>
<td>48</td>
<td>-24</td>
</tr>
<tr>
<td>Heschl's gyrus (BA 41)</td>
<td>48</td>
<td>-32</td>
</tr>
<tr>
<td>PHG (BA 28)</td>
<td>16</td>
<td>-10</td>
</tr>
<tr>
<td>MTG (BA 21)</td>
<td>44</td>
<td>-4</td>
</tr>
<tr>
<td>Subcallosal gyrus (BA 34)</td>
<td>26</td>
<td>6</td>
</tr>
<tr>
<td><strong>Parietal Lobe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMG (BA 40/7)</td>
<td>36</td>
<td>-54</td>
</tr>
<tr>
<td><strong>Subcortical</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claustrum</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>18</td>
<td>-64</td>
</tr>
<tr>
<td><strong>Left Hemisphere</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Frontal Lobe</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG pars orbitalis (BA 47)</td>
<td>-40</td>
<td>22</td>
</tr>
<tr>
<td>IFG pars triangularis (BA 45)</td>
<td>-46</td>
<td>22</td>
</tr>
<tr>
<td>Anterior insula (BA 13)</td>
<td>-32</td>
<td>22</td>
</tr>
<tr>
<td>Region</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td>-40</td>
<td>6</td>
</tr>
<tr>
<td>IFG pars opercularis (BA 44)</td>
<td>-42</td>
<td>2</td>
</tr>
<tr>
<td>Temporal Lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>aSTG (BA 22)</td>
<td>-50</td>
<td>10</td>
</tr>
<tr>
<td>pSTG (BA 22)</td>
<td></td>
<td>-46</td>
</tr>
<tr>
<td>Parietal Lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMG (BA 40)</td>
<td>-30</td>
<td>-50</td>
</tr>
<tr>
<td>SMG (BA 40/7)</td>
<td>-46</td>
<td>-56</td>
</tr>
<tr>
<td>Subcortical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amygdala</td>
<td>-18</td>
<td>-6</td>
</tr>
<tr>
<td>Caudate nucleus</td>
<td>-12</td>
<td>-4</td>
</tr>
<tr>
<td>Putamen</td>
<td>-22</td>
<td>14</td>
</tr>
<tr>
<td>Cerebellum</td>
<td></td>
<td>-28</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-8</td>
<td>-6</td>
</tr>
<tr>
<td>Midline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frntal Lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMA (BA 6)</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Occipital Lobe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuneus (BA 17)</td>
<td>0</td>
<td>-82</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>LP</th>
<th>U</th>
<th>AP</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Size (mm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Right Hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior Temporal Gyrus (BA 22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>799</td>
</tr>
<tr>
<td>Supramarginal gyrus (BA 40/7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>437</td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>115</td>
</tr>
<tr>
<td><strong>Left Hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supramarginal gyrus (BA 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Anterior insula (BA 13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>81</td>
</tr>
<tr>
<td><strong>Midline</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplementary motor area (BA 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>669</td>
</tr>
</tbody>
</table>
Table 3

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Size (mm^3)</th>
<th>AP Studies</th>
<th>LP Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AP &gt; LP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Right Hemisphere</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior temporal sulcus (BA 22)*</td>
<td>44</td>
<td>-40</td>
<td>0</td>
<td>75</td>
<td>9 (47%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>Middle temporal gyrus (BA 21)</td>
<td>44</td>
<td>-44</td>
<td>2</td>
<td>27</td>
<td>5 (26%)</td>
<td>1 (10%)</td>
</tr>
<tr>
<td><em>Left Hemisphere</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFG pars orbitalis (BA 47)</td>
<td>-40</td>
<td>28</td>
<td>-6</td>
<td>1063</td>
<td>9 (47%)</td>
<td>1 (10%)</td>
</tr>
<tr>
<td>Thalamus</td>
<td>-12</td>
<td>-6</td>
<td>16</td>
<td>197</td>
<td>3 (16%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td><strong>LP &gt; AP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Right Hemisphere</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td>50</td>
<td>8</td>
<td>30</td>
<td>2773</td>
<td>4 (21%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>Angular gyrus (BA 40)</td>
<td>50</td>
<td>-32</td>
<td>20</td>
<td>1437</td>
<td>4 (21%)</td>
<td>5 (50%)</td>
</tr>
<tr>
<td>Claustrum**</td>
<td>28</td>
<td>20</td>
<td>6</td>
<td>397</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>IFG pars opercularis (BA 44)</td>
<td>46</td>
<td>8</td>
<td>16</td>
<td>111</td>
<td>3 (16%)</td>
<td>9 (90%)</td>
</tr>
<tr>
<td><em>Left Hemisphere</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle frontal Gyrus (BA 9)</td>
<td>-42</td>
<td>8</td>
<td>30</td>
<td>321</td>
<td>4 (21%)</td>
<td>3 (30%)</td>
</tr>
<tr>
<td>IFG pars opercularis (BA 44)</td>
<td>-40</td>
<td>14</td>
<td>8</td>
<td>129</td>
<td>4 (21%)</td>
<td>4 (40%)</td>
</tr>
<tr>
<td>Superior temporal gyrus (22)</td>
<td>-50</td>
<td>-42</td>
<td>10</td>
<td>45</td>
<td>4 (21%)</td>
<td>4 (40%)</td>
</tr>
<tr>
<td><em>Midline</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuneus (BA 17)</td>
<td>0</td>
<td>-76</td>
<td>10</td>
<td>1349</td>
<td>4 (21%)</td>
<td>3 (30%)</td>
</tr>
</tbody>
</table>

* Peaks from either condition may be differentially localized as suggested by the left panel of figure 1.
**May be mislocalized from nearby Putamen and/or Insula
Table 4

<table>
<thead>
<tr>
<th>High&gt;Low Verbal Working Memory</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>ALE ($10^3$)</th>
<th>Same/Different Identification</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Right Hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Superior Temporal Gyrus (BA 22)</td>
<td>48</td>
<td>-30</td>
<td>12</td>
<td>7.80</td>
<td>5 (83%) 0 (0%)</td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td>52</td>
<td>12</td>
<td>32</td>
<td>6.98</td>
<td>4 (67%) 1 (25%)</td>
</tr>
<tr>
<td><strong>Left Hemisphere</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle frontal gyrus (BA 9)</td>
<td>-42</td>
<td>14</td>
<td>32</td>
<td>10.96</td>
<td>2 (33%) 0 (0%)</td>
</tr>
</tbody>
</table>