Effects of cortical and subcortical brain damage on the processing of emotional prosody

Marc D. Pell
School of Communication Sciences and Disorders
McGill University, Montreal, Canada
marc.pell@mcgill.ca

Abstract
Cortical and subcortical contributions to the processing of emotional speech prosody were evaluated by testing adults with single focal lesions involving the right hemisphere \(n=9\), adults with basal ganglia damage in idiopathic Parkinson’s disease \(n=21\), and healthy aging adults \(n=33\). Participants listened to semantically-anomalous utterances in two conditions (identification, rating) which assessed their recognition of five prosodic emotions. Findings confirmed that both right hemisphere and basal ganglia pathology were associated with impaired comprehension of prosody, although possibly for distinct reasons: right hemisphere compromise produced a more pervasive insensitivity to emotive features of prosodic stimuli, whereas basal ganglia disease produced a milder and more quantitative impairment on these tasks. The implications of these findings for differentiating cortical and subcortical mechanisms involved in prosody processing are considered.

1. Introduction
The right cerebral hemisphere in humans has long been ascribed a critical role in the communication of emotions, including how vocal-prosodic expressions of emotion are understood through relative changes in the pitch, loudness, and rhythmic structure of spoken language. Although it was initially believed that right cortical regions are exclusively specialized for processing emotional prosody in the brain, this idea has been revised in recent years in favour of the view that the right hemisphere serves an essential but shared role in this form of processing with functionally-related cortical and subcortical structures [1,2]. In particular, the importance of the basal ganglia to the processing of emotional prosody has been highlighted by a number of recent studies [3,4,5].

In the cognitive regulation of vocal expressions of emotion, it appears likely that cortical and subcortical structures furnish unique mechanisms which contribute to how the emotive contents of speech are processed and understood [2,3,5]. However, there have been few attempts to directly compare the processing of emotional prosody following cortical damage to the right hemisphere versus damage to the basal ganglia on a comprehensive and unitary set of tasks. Such comparisons may shed light on important quantitative as well as qualitative differences in emotional processing associated with cortical versus subcortical pathology, with implications for understanding how these different brain regions are involved at different stages of prosodic processing and/or in different ways.

Following two separate reports which examined prosodic processing in patients with right hemisphere damage (RHD) and in patients with basal ganglia disease in the early stages of idiopathic Parkinson’s disease (PD) [2,5], the present investigation undertook further analyses to characterize potential differences in how these two groups interpret the emotional significance of prosody from speech. More specifically, the ability of cortical and subcortical patients to interpret five ‘primary’ emotions was evaluated in two distinct processing conditions: one which required them to categorize the meaning of prosodic expressions from a closed set of alternatives; and one which examined how well patients detected the presence of these emotional meanings from prosody in a more graded manner using a six-point linear scale of “increased presence” of the evaluated emotion.

2. Methods

2.1. Subjects
Three groups of right-handed, English-speaking adults were studied. One participant group had suffered a single, thromboembolic event resulting in lesion of the right hemisphere of the brain \(n=9\). A second group had been diagnosed with idiopathic Parkinson’s disease without dementia in the mild to moderate range of motor severity \(n=21\). A third group was composed of healthy, normally-aging adults without neurological damage \(n=33\). Participants in the three groups were roughly comparable in age (HC: \(M = 64.0 \pm 9.0\); RHD: \(M = 64.2 \pm 16.6\); PD: \(M = 61.7 \pm 8.6\)), in the ratio of male to female participants (HC: 17:16; RHD: 4:5, PD: 11:10), and had all completed at least ten years of formal education (HC: \(M = 15.1 \pm 2.5\); RHD: \(M = 11.9 \pm 1.5\); PD: \(M = 16.0 \pm 3.7\)). All patient and control subjects passed a pure-tone audiometric screening prior to the study to ensure acceptable hearing thresholds (minimum 35 dB HL at 0.5, 1, 2 kHz, in the better ear) as well as a mental status exam (Mattis Dementia Rating Scale).

2.2. Stimuli
Prosody stimuli were short “pseudo-utterances” (e.g., “Someone migged the pazing”) or semantically-meaningful utterances (e.g., “I didn’t make the team”) produced by four male and four female actors. These stimuli were selected from a published inventory of recordings [6] in which these utterance types had been emotionally inflected to convey one of five target emotions—anger, disgust, happiness, sadness, and surprise. Selected items were perceptually validated to ensure that the intended emotion could be identified by at least 70% of a group of naïve listeners. The presentation of pseudo-utterances devoid of meaningful verbal-semantic
information ensured that all relevant cues to emotion were confined to the prosodic channel when these were presented. In contrast, semantically-meaningful utterances contained congruent prosodic and semantic cues which could be harnessed by listeners to bias an emotional interpretation when these stimuli were presented.

2.3. Tasks

Two task conditions varied how brain-damaged listeners judged the emotional significance of prosody in speech. Both conditions employed a common set of the pseudo-utterances representing each of the five target emotions, and one condition also presented semantically-meaningful utterances to evaluate whether semantic content helped to facilitate emotional judgements for either group.

(a) Emotion identification – Two tasks required participants to categorize the emotional meaning expressed using a multiple-choice response format with a closed set of the five target emotions. In the first task ("Pure prosody" identification), participants listened to a series of pseudo-utterances and then chose a verbal label corresponding to the emotion of the prosody. A second task ("Prosody-semantic" identification) required the same participants to name the emotion from the utterances which biased an emotional interpretation based on the prosody as well as the semantic content of the sentence. Each task was comprised of 40 trials (5 emotions x 8 items) and the accuracy of the response was recorded by a computer. Listeners were instructed to select the verbal label that "best describes the emotion of the speaker".

(b) Emotion rating – This task required the participants to judge the emotional significance of the 40 utterances presented in the identification task along a six-point scale indicating the amount of the emotion detected. Prosodic stimuli (pseudo-utterances) were presented on five separate occasions in which participants were required to rate each of the stimuli in reference to only one of the five target emotions (e.g., subjects rated each of the 40 items for how "happy" the speaker sounded, etc.). Participants were instructed to judge "how much" of the pre-selected target emotion was represented by each stimulus on a scale from 0 ("not at all") to 5 ("very much"). This process was repeated until each item had been evaluated in reference to all five target emotions. The distribution of numerical ratings assigned to stimuli when the listener was actual judging the intended target emotion (i.e., ratings for the 8 happy sentences when asked to judge these stimuli for their relative "happiness", etc.) was then examined and compared across the three groups.

2.4. Procedure

Participants were tested individually in a quiet room over a minimum of two sessions lasting one hour each. Tasks within each condition were randomized as much as possible and distributed across sessions to mitigate effects of stimulus repetition and fatigue. Emotional stimuli were presented over headphones from a portable computer, and responses were recorded directly by a response pad. No time limitations were imposed in any of the tasks, and frequent breaks were given during the testing session. All tasks were preceded by detailed instructions and practice.

3. Results

3.1. Emotion identification

The ability to identify the emotional meaning of utterances from underlying prosodic cues (pure prosody identification) or from combined prosody and semantic information (prosody-semantic identification) was assessed in a single 3 x 2 x 5 mixed design ANOVA. This analysis was characterized by a between-subjects factor of GROUP (HC, RHD, PD) and repeated factors of both CUES (Pure prosody, prosody-semantic) and EMOTION (anger, disgust, happy, sad, surprise).

The ANOVA yielded significant main effects for GROUP \(F(2,60) = 20.59, p<0.001\), CUES \(F(1,60) = 80.71, p<0.001\) and EMOTION \(F(4,240) = 12.07, p<0.001\), and an interaction of CUES x EMOTION \(F(4,240) = 3.94, p<0.01\). Of greatest interest here, the main effects were modulated by significant interactions of GROUP x CUES \(F(2,60) = 4.05, p=0.02\) and GROUP x EMOTION \(F(8,240) = 2.14, p=0.04\). Tukey’s posthoc elaboration of the Group by Emotion interaction revealed that the RHD group was less accurate than the other two groups in identifying all emotional targets with the exception of "surprise" (for which none of the groups differed). Overall, the PD group did not show selective deficits in identifying any of the five emotions relative to control subjects. Whereas the RHD patients varied little in how well they identified the five emotion types, both the HC and PD groups experienced significantly greater difficulty identifying "surprise" followed by "anger"; and both of these groups were most accurate in categorizing "sad" prosody.

Posthoc inspection of the Group x Cues interaction established that all three groups identified emotions more accurately when provided both prosody and verbal-semantic cues in speech ("prosody-semantic" task). When required to interpret emotion from prosody alone ("pure prosody" task), both patient groups were significantly impaired relative to the healthy control subjects, and the RHD patients performed significantly worse than the PD patients in this context. However, when subjects could also use the semantic content of sentences to arrive at emotional meanings, only the RHD group was significantly impaired relative to the control subjects. The influence of available emotion cues on group performance is illustrated by Figure 1.

![Figure 1: Group accuracy for emotion identification based on prosodic cues only ("pure-prosody" task) or based on prosody and semantic cues ("prosody-semantic" task).](image-url)
3.2. Emotion rating

The ability to rate the presence of pre-determined emotional qualities from prosodic stimuli was analyzed by focusing on the subset of ratings collected when subjects judged the intended emotional meaning of the stimulus (e.g., those ratings on the six-point scale that represented how “happy” a speaker sounded when the actual target of the item was happiness, and so forth). If target emotions were correctly recognized, a high frequency of ratings at the upper end of the ordinal scale was expected.

The frequency distribution of responses assigned at each interval of the rating scale was compared between groups using Chi square analyses summed across emotions (including 1354, 369 and 860 observations for the HC, RHD and PD groups, respectively). When compared to performance of the healthy adults, the pattern of ratings assigned to emotional prosody was significantly different overall in adults with RHD \( \chi^2(5) = 95.73, p<0.001 \) and PD \( \chi^2(5) = 33.39, p<0.001 \). The distribution of responses assigned by RHD versus PD patients was also statistically independent \( \chi^2(5) = 52.63, p<0.001 \), as illustrated for the three groups in Figure 2.

Qualitative inspection of group response tendencies implies that, whereas healthy adults show the expected trend for assigning an increasing number of higher ratings to prosodic stimuli when judging the actual target meanings present, the RHD patients exhibited a distinct pattern characterized by very little differentiation of rating responses to target meanings except at the highest interval of the ordinal scale (“5”), where the proportional frequency of their responses to target emotions resembled that of the HC participants. The PD patients’ ratings were more qualitatively similar to those of the HC group, but exhibited a notable reduction in the proportion of responses recorded at the upper end of the ordinal scale (i.e., “4” and especially “5” ratings).

![Figure 2: Detection of target emotions by the three groups from prosodic cues alone in reference to a six-point scale of “increased presence” of the perceived emotion. Data represent the percentage of total responses assigned at each interval of the scale, per group.](image)

4. Discussion

As previously demonstrated [2,3,5], the ability to understand emotional meanings from prosodic attributes of speech is impaired in patients with focal right hemisphere (cortical) lesions as well as patients with basal ganglia disease. This evidence was derived from two distinct evaluative contexts which required participants to engage in meaningful processing of prosodic expressions to either categorize or rate the presence of target meanings. Although neither patient group performed normally on receptive prosody tasks, present analyses underscore important quantitative and qualitative differences in the performance of individuals with cortical versus subcortical pathology. It was found that RHD adults exhibited pronounced deficits in their sensitivity to prosodic meanings which varied qualitatively from the other two groups, and which generalized to all of the emotions tested, whereas adults with PD were significantly but more lightly impaired in their recognition of emotion from prosody, and showed many qualitative similarities in their error patterns to the healthy control group which must be taken into account.

4.1 Right hemisphere contributions to emotional prosody

For the RHD group, the ability to identify the five emotions from prosody was found to be deficient in the face of prosodic cues along (“pure prosody” task) and when semantic information further biased the correct response, although adding semantic cues served to greatly facilitate emotion identification in the RHD group. It was also noted that emotion ratings gathered from members of the RHD group were highly distinct from those of the HC and PD groups, showing virtually no differentiation in the relative frequency of responses at the five lowest intervals of the six-point emotion scale (see [2] for a more detailed explanation). This pattern of ratings—which conflicted with the expected trend for assigning an increasing number of high (i.e., “3” to “5”) ratings to the target emotions known to be present in the signal—furnishes strong indications that the RHD patients failed to globally appreciate relative differences in the presence/intensity of prosodic expressions of emotion [2, 7].

Although further data will be needed to determine the precise source of prosodic deficits exhibited by RHD patients, current findings support research that stipulates a mandatory right hemisphere role in specialized functions for decoding the emotive significance of prosodic information in speech [1, 7]. Additional research will be needed to elucidate whether right hemisphere cortical lesions affect procedures which undertake early perceptual analyses of prosodic/pitch contours [8] and/or which map these features onto acquired knowledge or associations which specify their emotional-symbolic function in speech [9].

4.2 Basal ganglia contributions to emotional prosody

Whereas regions of the right hemisphere are critical to assign relevance to emotional prosody at the cortical level, the notion that understanding emotional prosody relies on subcortical mechanisms such as the basal ganglia is highly indicated by these and other data [3,4,5]. Here, individuals with PD were impaired in the ability to identify and rate the five emotions based on prosodic cues alone, although there was clear evidence in the identification condition that PD participants could successfully use semantic content to arrive at a correct
emotional interpretation; these findings imply that the effects of basal ganglia disease are more specific to processing aspects of prosodic structure rather than processing emotional information more generally. The observation that deficits in the PD group were less severe than those of the RHD group in both prosody conditions [3], and that their error patterns showed many qualitative similarities among the five emotions to those of healthy listeners, reinforces the view that basal ganglia disease affects prosody in a notable manner, but that it does not disrupt the retrieval of emotional meaning from prosodic cues to the same extent as right hemisphere cortical lesions.

As suggested recently [5,10], the decoding of emotional prosody may be susceptible to basal ganglia disease because these structures supply a critical cognitive mechanism which promotes the temporal binding of prosodic cue sequences during on-line speech processing. The basal ganglia are known to play a central role in time perception and interval/sequence judging [11], and by extension, it is plausible that a functionally-related mechanism for facilitating the coherence of meaningful prosodic features over time is well suited to underlying capabilities of the intact basal ganglia (striatum). According to this view, the basal ganglia would be necessary to refine the prosodic input by “potentiating” vocal cue sequences with implicit representational value in speech, thereby facilitating successful mapping of prosodic feature constellations onto right hemisphere cortical associations which then retrieve their emotive significance during communication [5,12].

6. Conclusions

If one compares the broader data on patients with RHD and those with basal ganglia pathology, one can speculate that the basal ganglia, via the striatal-pallido-thalamocortical frontal “cognitive” loops, act as an intermediary to (relatively) right-sided cortical operations which extract long-term pitch variations of the auditory signal and those which subsequently activate knowledge about the emotional significance of prosodic cues. This series of operations would facilitate a cognitively-elaborated response about the emotive value of prosodic information in speech in tandem with various forms of (relatively) left-hemisphere dominant linguistic operations which are simultaneously engaged by the speech input [13]. As representations about the emotive qualities of prosody become activated in memory, one might anticipate certain “convergence points” in linguistic-semantic processing which promote comparison and consolidation of emotionally-relevant activations contained in the two speech channels; these procedures might influence the nature of initial assumptions drawn about the prosody, the semantic content, or both, promoting further analysis in both channels.

This dynamic interplay of cortical and subcortical operations, while strictly hypothetical at present and in need of empirical verification, would begin to explain how listeners adopt a unitary and contextually-appropriate impression of a speaker’s emotive condition, their attitudinal stance, and other interpersonal information that is routinely derived from speech.

7. Acknowledgements

I am grateful to Elmira Chan for help in data collection and organization, and to the Canadian Institutes of Health Research for financial support.

8. References