Motivation and Emotion
https://doi.org/10.1007/s11031-025-10147-9

ORIGINAL PAPER

The interplay between emotional semantics and prosody: behavioural and skin conductance responses

Francesca Carbone[1,2] · Piera Filippi[3,4] · Caterina Petrone[2]

Accepted: 1 July 2025

Abstract

Listeners recover speakers' emotional state by integrating information from emotional semantics and the prosody accompanying sentence production. However, little attention has been paid to how the processing of emotional semantics and prosody is mediated by emotional dimensions (arousal and valence) and whether this relationship is mirrored at the level of the autonomic nervous system. In the present study, we compared the effects of emotional semantics, emotional prosody and their combination on behavioural responses and skin conductance responses (SCRs). Native French listeners ($N=77$) judged the arousal and valence of utterances conveying emotions through semantics and/or prosody, while their SCRs were recorded. Our behavioural findings showed that prosody was more effective than semantics in transmitting emotional information to the listeners. Regarding SCRs, we found a significant effect of angry prosody in female listeners only, independently of whether it was combined with emotional or neutral semantics. Our work provides evidence for the perceptual saliency of emotional prosody and suggests a sex-based differentiation in emotional speech processing at the autonomic level. Finally, our findings align with the idea of emotional speech processing as a multistage process involving various processing levels, some of which may be driven by the autonomic nervous system and mediated by specific factors such as the emotional dimensions of stimuli and the sex of listeners.

Keywords Emotional prosody · Emotional semantics · Emotion processing · Peripheral responses

Introduction

In everyday life, listeners infer speakers' emotions by integrating multiple emotional cues. For instance, they rely on the co-occurrent stream of semantic content (words or sentences with emotional meaning, e.g., 'I hate you', 'I won the lottery') and prosodic features (e.g., increase in fundamental frequency, speech rate and intensity) to identify the speaker's emotions. Integrating emotional semantics and prosody is essential for accurate emotion recognition (e.g., Kotz & Paulmann, 2007; *inter alia*). Behavioural studies consistently show that congruent combinations of emotional semantics and prosody enhance both the speed and accuracy of emotion recognition compared to when each cue is presented in isolation (Paulmann & Pell, 2011; Kotz & Paulmann, 2011; Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020). Similarly, ERP studies have revealed a neural processing advantage for congruent combinations of emotional semantics and prosody over single cues (Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Paulmann et al., 2009).

In addition, past research has investigated the relative contribution of semantic and prosodic cues to emotion processing (Paulmann & Pell, 2011; Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020, 2022). A growing body of research suggests that prosody may override semantics during emotion processing (Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020, 2022). In an identification Stroop task (Filippi et al., 2017a), listeners were presented with incongruent combinations of emotional semantics and prosody (e.g., a positively valenced word spoken with a negatively valenced prosody, such as the word *"Happy"*

✉ Francesca Carbone
f.carbone@kent.ac.uk

1 School of Psychology, University of Kent, Canterbury, UK

2 Aix-Marseille Univ, CNRS, LPL, Aix-en- Provence, France

3 Department of Philosophy, University of Zurich, Zurich, Switzerland

4 Institute for the Interdisciplinary Study of Language Evolution, University of Zurich, Zurich, Switzerland

Published online: 20 August 2025

spoken with sad prosody). Accuracy was higher when listeners focused on prosody (while ignoring semantics) than when focusing on semantics (while ignoring prosody), suggesting difficulty in disregarding prosodic cues. The authors interpreted these findings as evidence of the perceptual salience of emotional prosody, which may capture listeners' attention more effectively than emotional semantics (Filippi et al., 2017a). These results are consistent with studies showing that listeners can reliably infer speakers' emotions from prosody alone, even in the absence of meaningful semantic content (e.g., in pseudo-utterances, Banse & Scherer, 1996; Johnstone & Scherer, 2000; Scherer, 2003; Roche et al., 2015). However, these findings contrast with evidence suggesting that the relative importance given to emotional semantics and prosody varies across cultural and linguistic backgrounds. For instance, studies have shown that American English listeners, who come from a low-context culture (i.e., cultures where communication relies heavily on explicit, verbal information), tend to show an attentional bias toward emotional semantics. In contrast, Japanese listeners, from a high-context culture (where meaning is more often derived from contextual and nonverbal cues), show a greater sensitivity to emotional prosody (Kitayama & Ishii, 2002; Ishii et al., 2003).

The above-mentioned studies offer insights into the combined or isolated effects of emotional semantics and prosody on listeners' brain activity (e.g., ERP measures) and behavioural responses (e.g., recognition tasks). Building upon these studies, Schirmer and Kotz (2006) proposed a multistage model of emotional speech processing. The first stage involves sensory analysis of the speech signal and extraction of emotionally relevant acoustic cues at the auditory cortex level. In the second stage, the emotionally significant information is derived from acoustic and semantic cues via a pathway running from the superior temporal gyrus to the anterior superior temporal sulcus. In the final stage, this emotional information is made available to higher-order cognitive processes. For example, explicit evaluative judgments of emotional prosody are mediated by the right inferior frontal gyrus and orbitofrontal cortex. However, as Schirmer and Kotz (2006) noted, this model focuses only on cortical and subcortical mechanisms and overlooks the role of the autonomic nervous system (ANS).

The ANS has typically been considered a key component of emotion regulation and processing in research on emotional communication in both animals and humans (Levenson, 2003). In particular, the sympathetic branch of the ANS is responsible for many automatic, non-voluntary responses that are essential for preserving the well-being of individuals and species (Kreibig, 2010). These responses support rapid adaptation and prepare the body for different types of actions, such as fighting, freezing, or fleeing when facing fear or anger, or approaching a desired or pleasant stimulus (Panksepp, 1982). Sympathetic responses are triggered by emotional stimuli varying along both the arousal and valence dimensions. Studies in the visual domain have shown a positive correlation between subjective ratings of arousal and the ANS, while findings on the relationship between valence and ANS responses have been mixed (Burriss et al., 2007; Gomez & Danuser, 2010; *inter alia*).

Because of its older phylogenetic origins, emotional prosody may be biologically more relevant and thus more effective than emotional semantics in triggering sympathetic responses. Indeed, components of emotional prosody have been identified in the vocal communication systems of non-human mammals (Zimmermann et al., 2013; Briefer, 2012; Fischer, 2021), including non-human primates (Kershenbaum et al., 2016; Fischer & Price, 2017). Previous studies have hypothesized that prelinguistic hominids likely used the prosodic features of calls from their own and other species to detect emotional information relevant for survival, such as the presence of danger or predators (Falk, 2004; Belin et al., 2008; Zimmermann et al., 2013; Filippi et al., 2017a). Consequently, they may have developed rapid adaptive responses, such as fight, freeze, or flight reactions, to emotional signals like danger or alertness expressed through voice modulation, prior to the emergence of verbal communication systems (Darwin, 1871; Nesse, 1990; Brown, 2017; Filippi, 2020).

A technique that provides accessible indices of autonomic nervous system activity is electrodermal activity. Electrodermal activity (EDA) is an umbrella term referring to autonomic changes in the electrical properties of the skin. Its most widely studied property is skin conductance, which includes rapid phasic components (Skin Conductance Responses, SCRs) that result from sympathetic neuronal activity (Dawson et al., 2000; Boucsein, 2012). To our knowledge, only a few studies have investigated the isolated effect of emotional semantics and prosody on SCRs. Harris et al. (2003) showed that highly arousing, unpleasant words spoken with neutral prosody (i.e., reprimands such as 'Shut-up') elicited larger SCRs than less arousing unpleasant words (e.g., 'Cancer'), pleasant (e.g., 'Bride') and neutral words (e.g., 'Table'). Similarly, Aue et al. (2011) investigated the effect of pseudo-words (e.g., "goster") spoken with hot angry or neutral prosody on SCRs. They found that angry prosody, generally characterised by high arousal and negative valence (Ekman, 1999; Scherer, 2003), elicited larger SCRs than neutral prosody. These results suggest that emotionally arousing, unpleasant words and prosody have a strong impact on SCRs. They also support the 'negativity bias' hypothesis, which posits that listeners are more responsive to unpleasant stimuli than to pleasant or neutral ones. This bias may stem from the greater relevance of negative stimuli for survival, as they may serve as early warnings for potential threats and help the body prepare for rapid action (Cacioppo & Berntson, 1999; Aue et al., 2011).

To date, experimental evidence on the impact of emotional semantics and prosody on the sympathetic nervous system, particularly as measured by SCRs, is limited, leaving the following gaps for investigation. First, none of the above studies examined the relationship between subjective ratings of emotional dimensions (arousal and valence) and SCRs in response to emotional semantics or prosody. Second, a major limitation of the current literature is its focus on the isolated effects of individual cues, which have not been directly compared at the autonomic level. Third, previous research has concentrated on narrow subcategories of words (e.g., reprimands; Harris et al., 2003), neglecting the effects of broader emotional categories (angry vs. happy words or sentences). Regarding prosody, previous studies have been limited to hot anger (Aue et al., 2011), while the effects of pleasant, arousing prosodies (e.g., happy prosody) on SCRs remain unexamined. This gap is particularly relevant because, although the influence of arousal on SCRs is well established, empirical findings on valence-related differences remain inconsistent (Burriss et al., 2007; Gomez & Danuser, 2010; *inter alia*).

In the present study, we tested, for the first time, the impact of emotional semantics, emotional prosody and their combination on behavioural and SCRs. Listeners rated each stimulus on two emotional dimensions: arousal (i.e., the intensity of physiological activation triggered by the emotion) and valence (i.e., the degree of unpleasantness or pleasantness). At the same time, SCRs were recorded as an index of sympathetic activation. We focused on two basic and evolutionarily significant emotions, anger and happiness, expressed through semantics and/or prosody. These emotions are opposite in valence yet comparable in arousal, as both are generally considered high-arousal emotions (Ekman, 1999). They play a key role in adaptive behaviours and survival (Panksepp, 1982). Anger, a high-arousal negative emotion, is associated with a range of adaptive responses to threat, including fight, flight, or freezing reactions, depending on context and appraisal (Harmon-Jones, 2003). Happiness, in contrast, is a high-arousal positive emotion often linked to reward, social bonding, and approach behaviour (Berridge & Kringelbach, 2015). These differences in emotional valence and motivational direction may result in distinct autonomic profiles. For instance, anger-related stimuli might elicit larger SCRs, reflecting the evolutionary relevance of detecting threats for survival (Cacioppo & Berntson, 1999), whereas happy-related stimuli may trigger more moderate SCRs, reflecting approach motivation (Kreibig, 2010; Bradley & Lang, 2000). Furthermore, to assess the impact of emotional semantics and prosody, we used meaningful utterances expressing anger and happiness at the utterance level.

We expected listeners to assign more extreme arousal and valence ratings to stimuli conveying anger and happiness through both semantics and prosody than to those conveying the same emotions through a single cue (either semantics or prosody alone). This expectation aligns with behavioural studies suggesting that combined emotional cues offer a processing advantage (e.g., Kotz & Paulmann, 2011; Paulmann & Pell, 2011; Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020). Additionally, we predicted that stimuli conveying the emotions through prosody alone would receive more extreme emotional ratings than those conveying the emotions through semantics alone, consistent with evidence that emotional prosody is more perceptually salient than emotional semantics (Ben-David et al., 2016; Filippi et al., 2017b; Lin et al., 2020).

We also expected that sympathetic responses (SCRs) would partly reflect behavioural ratings of emotional dimensions. Specifically, we predicted that emotional prosody would elicit larger SCRs than emotional semantics, regardless of whether it was combined with emotional or neutral semantic content. This hypothesis is consistent with findings suggesting that emotional prosody may be more effective than emotional semantics in triggering adaptive responses via the autonomic nervous system (e.g., fight, freeze, or flight reactions; Aue et al., 2011), possibly due to its older phylogenetic roots (e.g., Zimmermann et al., 2013; Filippi et al., 2017b; Filippi, 2020). Additionally, we expected higher SCRs to unpleasant stimuli compared to pleasant and neutral ones, in line with the ‘negative bias’ hypothesis (Cacioppo & Berntson, 1999; Aue et al., 2011). Given the sex-based differences reported in the neurocognitive literature on emotion processing (e.g., Schirmer et al., 2002, 2005; Schirmer & Kotz, 2006), listener sex was included as a predictor in our analyses to explore whether behavioural responses and SCRs to emotional semantics and prosody varied accordingly, although investigating sex differences was not the primary focus of the present study.

Finally, we aimed to explore the relationship between listeners’ arousal and valence ratings and their SCRs. Consistent with our previous hypotheses, we predicted that larger SCRs would be associated with higher arousal ratings and lower valence ratings.

Method

Participants

Seventy-seven native French speakers aged between 18 and 35 years (41 females, mean age: 23.9, *SD*: 0.4) participated

Table 1 Sentences used in the EDA experiment, each spoken with either prosody matching the emotional content (congruent) or with neutral prosody

Neutral meaning	Angry meaning	Happy meaning
Il a amené l'enveloppe (He brought the envelope)	*Elle l'a poussé à bout* (She pushed him to the limit)	*Il a eu sa promotion* (He got his promotion)
Il est allé en ville (He went into town)	*Il lui a fait une crise* (He lost his temper with her)	*Il a obtenu son diplôme* (He got his degree)
Elle a pris les clés (She took the keys)	*Elle a hurlé contre lui* (She shouted at him)	*Elle a trouvé l'amour* (She found love)
Elle a lu l'article (She read the article)	*Elle a joué avec ses nerfs* (She played with his nerves)	*Elle a fait des progrès* (She made progress)

in the main experiment. The study was non-invasive and conducted in full compliance with the Declaration of Helsinki. All participants provided written informed consent prior to the experiment and received a €10 voucher as compensation for their participation.

Materials

The final set of stimuli used in the main experiment is available upon request from the author. The datasets and analysis scripts for the validation tasks are publicly accessible at the following link: https://osf.io/n5683/.

The stimuli consisted of twenty-eight utterances that varied in both emotional semantics and prosody. The verbal content conveyed anger, happiness, or a neutral state through emotion-laden words that implied the target emotion without explicitly naming it (e.g., anger: *Elle a joué avec ses nerfs* 'She played with his nerves'; happiness: *Elle a trouvé l'amour* 'She found love'; neutral: *Il est allé en ville* 'He went downtown'). All sentences were comparable in length (ranging from six to ten syllables) and followed the same syntactic structure: a subject followed by a verb and a direct or indirect object. The subject was consistently a third-person singular pronoun (*Elle/Il* 'She/He'). The complete list of sentences is provided in Table 1.

Sentences were recorded by a professional actress and were selected through a series of validation tasks (see details in the following section). After the recordings, three conditions were obtained depending on the cue(s) through which the emotions were delivered: (1) 'semantics only' (SO): sentences with angry or happy verbal meaning, spoken with neutral prosody; (2) 'prosody only' (PO): sentences with neutral verbal meaning, spoken with angry or happy prosody; (3) 'semantics + prosody' (SP): sentences with angry, happy or neutral verbal meaning, spoken congruently with angry, happy or neutral prosody. Twenty-four of the twenty-eight utterances conveyed anger or happiness through semantics only, prosody only or the combination of semantics and prosody (4 sentences X 2 emotions X 3 cue conditions). The remaining four utterances, which were used as the neutral baseline, conveyed neutral meaning and were spoken with neutral prosody ('Baseline_neutral'). Table 2 illustrates the experimental conditions and the corresponding labels.

Table 2 Experimental conditions and labels adopted in this study. SO = semantics only, PO = prosody only, SP = semantics + prosody

Number of sentences	Cue(s)	Semantics	Prosody	Experimental conditions/ Labels
4	semantics + prosody	neutral	neutral	Baseline_neutral
4	semantics only	happy	neutral	SO_happy
4	prosody only	neutral	happy	PO_happy
4	semantics + prosody	happy	happy	SP_happy
4	semantics only	angry	neutral	SO_angry
4	prosody only	neutral	angry	PO_angry
4	semantics + prosody	angry	angry	SP_angry

Recordings

Recordings were conducted in the anechoic chamber at the *Laboratoire Parole et Langage* in Aix-en-Provence, France. A professional French actress recorded fifteen sentences—five conveying anger, five conveying happiness, and five conveying a neutral state through their verbal content. The fifteen sentences were selected from an initial set of thirty sentences based on prior validation of their semantic content (see below).

Sentences were produced with three types of emotional prosody: angry, happy, and neutral. The actress memorized each sentence beforehand and was then instructed to produce it in accordance with the target emotion indicated by the experimenter. The actress wore a high-quality head-mounted microphone to ensure consistency in loudness. Recordings were captured at a sampling rate of 48 kHz.

A total of thirty-five utterances were produced: ten with neutral verbal content spoken with either angry or happy prosody ('prosody only' condition); ten with angry or happy verbal content spoken with neutral prosody ('semantics only' condition); ten with angry or happy verbal content spoken with matching angry or happy prosody ('semantics + prosody' condition); and five with neutral verbal content spoken with neutral prosody ('Baseline_neutral' condition). Of these thirty-five, twenty-eight target utterances (see Tables 1 and 2) were selected based on a series of validation tasks conducted prior to the main study.

Validation tasks

The validation tasks were administered through two online surveys via Qualtrics. Participants recruited for the validation tasks did not participate in the main experiment. Stimuli were validated in both their written and oral forms to ensure that semantic content and prosody accurately represented each target emotion.

Prior to the recordings, fifty-nine native French speakers (52 women; mean age = 25.7 years, *SD* = 1.3) participated in the validation of the written semantic content of an initial set of thirty sentences. Participants were presented with the sentences in written form and asked to: (i) identify the conveyed emotion in a three-alternative forced-choice task (neutral, happy, or angry); (ii) rate how representative each sentence was of the selected emotion on a 5-point Likert scale (1 = "not at all representative," 5 = "very representative"); and (iii) rate the emotional valence of each sentence on a 5-point Likert scale (1 = "very unpleasant," 5 = "very pleasant"). To mitigate fatigue effects, participants were divided into two lists ($n = 28$ and $n = 31$), with each list evaluating fifteen sentences and completing all three tasks for each sentence. Only fifteen sentences with a correct identification rate above 80% were retained and used for the recordings, resulting in a set of thirty-five utterances varying in both emotional verbal content and prosody (see above for details about the recordings).

After the recordings were completed, a separate group of sixty native French speakers (36 women; mean age = 26.8 years, *SD* = 3.4) validated the prosody of the thirty-five utterances recorded by the actress. Participants listened to the recordings and performed the same tasks as in the semantic validation: (i) emotion identification, (ii) representativeness rating, and (iii) emotional valence rating. They were asked to focus solely on prosody ('tone of voice'), ignoring the verbal content of the utterances. Additionally, they evaluated the plausibility of each utterance—i.e., how natural and realistic the combination of semantics and prosody appeared—on a 5-point Likert scale (1 = "not at all plausible," 5 = "very plausible"). Participants were randomly assigned to one of three lists ($n = 20$ each): one completed the emotion identification and representativeness tasks, another the valence task, and the third the plausibility task. Seven utterances were excluded due to identification accuracy below 80%. The remaining twenty-eight utterances were retained for use in the main experiment (see Table 2).

Furthermore, the arousal level of the twenty-eight selected stimuli was assessed separately for both the semantics and prosody of the stimuli. Thirty native French speakers (15 women; mean age = 30.6 years, *SD* = 7.7) rated the arousal elicited by each sentence in written form, focusing solely on its semantic content. Another group of thirty participants (18 females; mean age = 33.2 years, *SD* = 11.07) rated the arousal conveyed by the recorded utterances, instructed to focus exclusively on prosody while disregarding semantic content. In both tasks, participants used a 5-point Likert scale (1 = "weakly arousing", 5 = "strongly arousing").

The twenty-eight selected stimuli showed high identification accuracy for both (written) semantics and prosody (semantics: $M = 95.2\%$, $SD = 0.2$; prosody: $M = 99.3\%$, $SD = 0.1$), along with high representativeness ratings (semantics: $M = 4.2$, $SD = 0.9$; prosody: $M = 4.3$, $SD = 0.8$), with no significant differences across emotion categories (e.g., happiness vs. anger) or prosodic conditions (e.g., 'prosody only' vs. 'semantics + prosody' conditions; all $p > 0.01$). As expected, valence ratings differed significantly across emotion categories for both semantics and prosody (all $p < 0.01$): happy stimuli were rated as the most pleasant (semantics: $M = 4.7$, $SD = 0.5$; prosody: $M = 4.1$, $SD = 0.9$), followed by neutral (semantics: $M = 3.1$, $SD = 0.4$; prosody: $M = 2.9$, $SD = 0.3$), and angry stimuli as the most unpleasant (semantics: $M = 1.8$, $SD = 0.7$; prosody: $M = 1.8$, $SD = 0.9$). Arousal ratings also varied significantly across emotion categories consistently for both semantics and prosody (all $p < 0.01$): angry stimuli were rated as the most arousing (semantics: $M = 4.1$, $SD = 0.8$; prosody: $M = 4.1$, $SD = 0.8$), followed by happy stimuli (semantics: $M = 3.4$, $SD = 1.2$; prosody: $M = 3.6$, $SD = 0.7$), and neutral stimuli (semantics: $M = 1.3$, $SD = 0.6$; prosody: $M = 1.4$, $SD = 0.7$). Arousal ratings did not differ significantly between semantics and prosody within each emotion category (all $p > 0.01$). Specifically, angry prosody showed arousal ratings comparable to those of angry semantics, with the same pattern observed for happy and neutral stimuli (all $p > 0.01$). Additionally, angry prosody received similar arousal ratings whether presented alone ('prosody only' condition) or combined with congruent semantic content ('semantics + prosody' condition), a pattern also found for happy prosody (all $p > 0.01$). Neutral prosody showed no significant differences in arousal when comparing the 'semantics only' and 'Baseline_neutral' conditions (all $p > 0.01$). No significant differences in plausibility were found across emotion and cue conditions (all $p > 0.01$).

Acoustic analyses

The acoustic analyses were conducted using PRAAT (Boersma, 2001). Three main prosodic parameters were analysed for each of the 28 utterances: intensity, fundamental frequency (f0) and utterance duration. Utterances spoken with angry and happy prosody showed higher f0 mean, broader f0 range, and greater intensity than utterances spoken with neutral prosody (all $p < 0.01$). No acoustic differences were found between the 'semantics + prosody' and 'prosody only' conditions within each emotion category (all $p>0.01$). Additionally, no acoustic differences were found between utterances spoken with angry and happy prosody ($p>0.01$). Furthermore, no differences were found between utterances with emotional verbal content spoken with neutral prosody ('semantics only' condition) and those with neutral content spoken with neutral prosody ('Baseline_neutral') (all $p>0.01$). No differences were found in duration across emotion and cue conditions (all $p>0.01$) (Table 3).

Table 3 Means and standard deviations (in parentheses) of the acoustic parameters for utterances produced with neutral, happy, and angry prosody

Prosody	Duration (sec)	f0 mean (Hz)	f0 range (max-min; Hz)	Intensity mean (dB)
neutral	1.09 (0.2)	188 (7.8)	99 (10.3)	71 (1.1)
happy	1.09 (0.1)	317 (9.9)	166 (34.2)	76 (2.1)
angry	1.10 (0.1)	306 (8.3)	175 (39.1)	77 (1.7)

Procedure

Participants were seated in a silent room at the *Laboratoire Parole et Langage* and listened to the stimuli through professional-grade headphones. They were instructed to attend to the entire utterance and evaluate the emotion conveyed by the speaker, taking into account both prosody (i.e., the 'tone of voice') and the verbal content. First, they rated the arousal of each stimulus using a 5-point Likert scale (1 = 'weakly arousing'; 5 = 'strongly arousing') by pressing a button box with their dominant hand. Next, they rated the valence of each stimulus on a separate 5-point Likert scale (1 = 'very unpleasant'; 5 = 'very pleasant'). Electrodermal activity was simultaneously recorded on the volar surface of the medial and distal phalanges of the fingers of the participants' non-dominant hand (Dawson et al., 2000).

Figure 1 illustrates the experimental paradigm. Each trial began with a fixation cross displayed for 2000 ms, followed by the auditory stimulus, which was played for 3000 ms. The arousal and valence scales were then presented on the screen, each shown separately in a fixed order—first the arousal scale, then the valence scale. Each scale remained visible for a minimum of 3500 ms and disappeared only after the participant responded by pressing a button. After a 3500-ms blank screen, the next trial began. The interstimulus interval (i.e., the time lapse between the presentation of an auditory stimulus and of the next one) was at least 15.5 s to account for the slow changes in electrodermal activity (skin conductance responses typically begin 1–3 s after stimulus onset, with a half recovery time between 2 and 10 s; Dawson et al., 2000). The stimuli were split into two

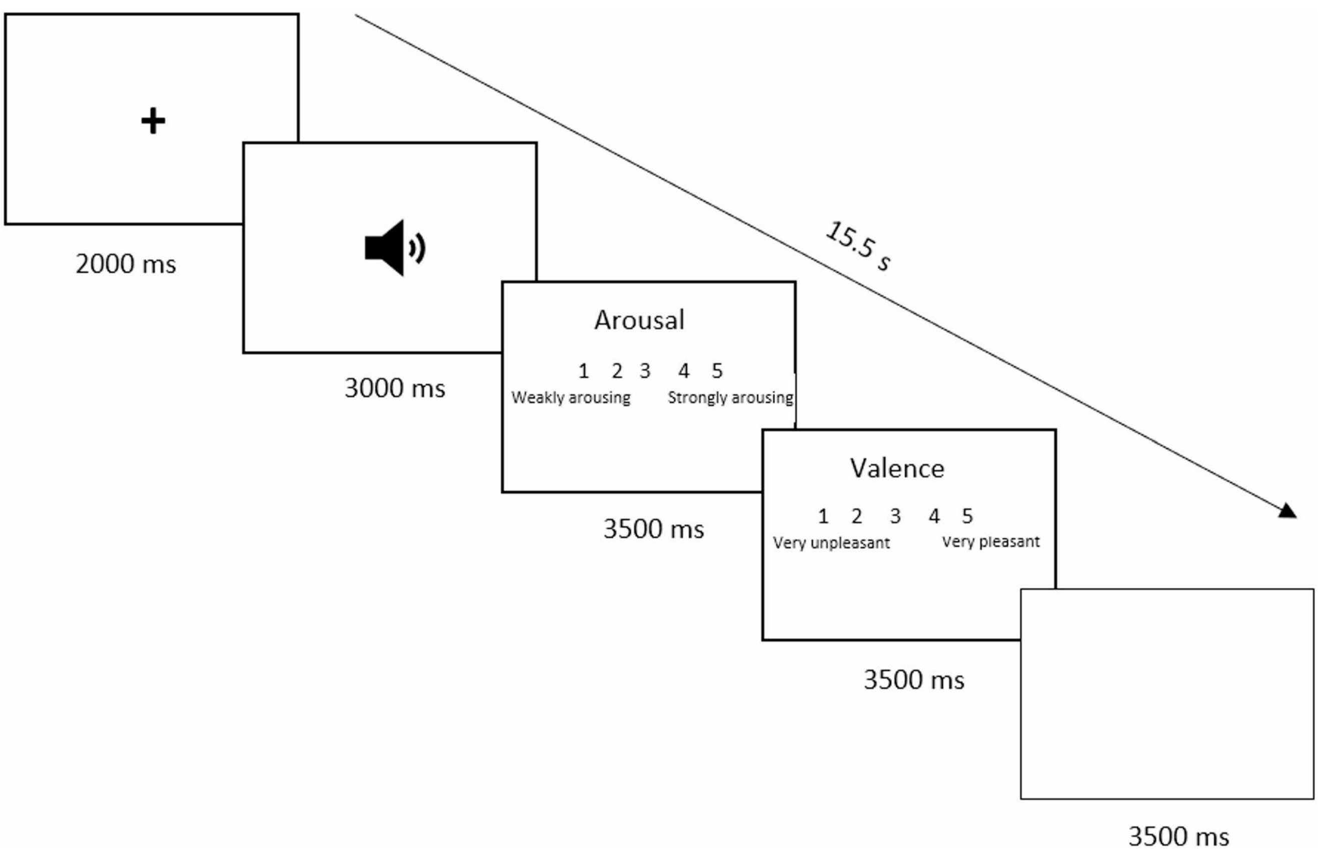

Fig. 1 Illustration of the experimental paradigm

blocks and presented once in each block, with a brief pause between them. Stimuli were randomized within each block, and block order was counterbalanced across participants. A short training session preceded the experiment in which participants were asked to rate three items (which were not included in the experimental session) on the arousal and valence scales. The total duration of the experiment was kept short (around 10 min) to avoid habituation effects.

Settings and apparatus

The auditory stimuli were presented via *E-prime 2.0* (Psychology Software Tools, Inc., Pittsburgh, PA, USA). Electrodermal activity was simultaneously recorded using a BIOPAC M36R system and *Acqknowledge 4.4* software (Biopac Systems, Inc., Goleta, CA, USA). Before starting the experiment, two snap Ag–AgCl electrodes (11 mm diameter active area), pre-gelled with isotonic gel, were attached to the volar surface of the medial and distal phalanges of the fingers of the participants' non-dominant hand (Dawson et al., 2000). Listeners were instructed to place their non-dominant hand palm-up on a table and to minimise movement as much as possible throughout the experiment, in order to reduce potential motion artefacts from hand movements.

Recordings, filtering and analyses followed standard procedures. The raw signal was converted into a waveform with a time resolution of 100 Hz, filtered using a high-pass filter with a cutoff frequency of 0.05 Hz and down sampled to a 10 Hz sampling rate (Boucsein, 2012).

The outcome variable was the SCR amplitude for each stimulus, calculated from the difference between the baseline (level of skin conductance over the two seconds before stimulus onset) and the peak SCR following stimulus presentation. These values were extracted from the temporal window spanning 1–3 s after stimulus onset. The minimum threshold for SCR detection was set at 0.01 µS. In case of multiple peaks over 0.01 µS, the values were averaged prior to statistical analysis. Six participants were excluded from the statistical analyses due to non-responsiveness to all stimuli (Herrero et al., 2020).

Data analysis

The full datasets and analysis scripts are available on the Open Science Framework webpage (see https://osf.io/n5683/). Statistical analyses were performed in the R environment (R Development Core Team, 2012, version 3.5.1). Graphical exploration was performed using the *ggplot2* package (version 3.4.0; Wickham, 2016).

A series of linear mixed models was conducted using the *lme4* package (version 1.1–31; Bates et al., 2015). The following dependent variables were entered in separate analyses: (i) ratings of arousal; (ii) ratings of valence; (iii) SCR amplitude. Prior to the statistical analysis, SCR amplitude was transformed using the square root transformation to account for individual differences (Dawson et al., 2000). Arousal and valence were also log-transformed to achieve a normal distribution. Data from the behavioural ratings, though, were plotted without logarithmic transformation to display them on the original scales.

For each model, we tested the fixed factors, Emotional cue (seven levels: Baseline_neutral, SO_happy, PO_happy, SP_happy, SO_angry, PO_angry, SP_angry; see Table 1) and Listener sex (two levels: women, men), and their interaction (Listener sex * Emotional cue). The Emotional cue factor was created by combining the Cue (three levels: SO: semantics only; PO: prosody only, SP: semantics + prosody) and Emotion factors (two levels: angry, happy), plus the Baseline_neutral. This combination was necessary due to the partial independence of the Cue and Emotion factors. Specifically, the Emotion factor was nested within the Cue factor (e.g., both anger and happiness are conveyed through semantics only, prosody only and the combination of semantics and prosody). Furthermore, the two factors shared the common baseline (Baseline_neutral) corresponding to stimuli with neutral emotional semantics spoken with neutral prosody. Random intercepts for Listener and Sentence were also included, along with by-listener random slopes for Emotional cue.

The Baseline_neutral level was entered as the intercept of the model. For pairwise comparisons involving levels other than neutral (happy stimuli vs. angry stimuli), the intercept was changed, and the p-values were adjusted through the Bonferroni correction. Only p-values smaller than 0.017 were considered significant, with a significance threshold adjusted for multiple comparisons ($p = 0.05$ divided by the number of models (3) run for each dependent variable).

The model for (1) arousal ratings, (2) valence ratings and (3) SCR amplitude) was:

Dependent Variable (Arousal or Valence or SCR amplitude) ~ Emotional cue * Listener sex + (1 + Emotional cue| Listener) + (1| Sentence).

When convergence issues arose, the random structure of the models was simplified to reduce overparameterization—for example, by replacing random slopes with random intercepts only or by removing random components with very little variance. This simplification was applied only to affected models and did not change the interpretation of the results.

Finally, a series of non-parametric correlations (Kendall's Rank Correlation) was performed to examine the relationship between participants' behavioural responses (rating of arousal and valence scores from the main experiment) and SCRs. Kendall's Rank Correlation is particularly suited for the analysis of ordinal data (e.g., ratings from Likert scales), and that does not fit a normal distribution (Hollander et al., 2013). Given the number of variables included in the correlation, the alpha level for significance was set to 0.01.

Results

Table 4 shows the descriptive statistics and p-values for the main variables and conditions. Results are described below separately for the behavioural ratings (arousal and valence), SCRs and the correlation analysis.

Behavioural responses

Figure 2 illustrates the results of the arousal and valence ratings as a function of emotional cue and listener sex. Arousal ratings varied across cue conditions. Angry and happy stimuli were judged as more arousing than neutral ones when the emotion was conveyed either by the combination of semantics and prosody (SP_angry vs. Baseline_neutral: [$\beta = 0.95$, SE = 0.09, $t = 10.94$, $p < 0.001$]; SP_happy vs. Baseline_neutral: [$\beta = 0.99$, SE = 0.07, $t = 15.08$, $p < 0.001$]) or by prosody only (PO_angry vs. Baseline_neutral: [$\beta = 0.93$, SE = 0.08, $t = 10.93$, $p < 0.001$]; PO_happy vs. Baseline_neutral: [$\beta = 0.92$, SE = 0.07, $t = 13.96$, $p < 0.001$]). No significant differences were found between the 'semantics + prosody' and 'prosody only' conditions (SP_angry vs. PO_angry: [$\beta = 0.03$, SE = 0.04, $t = 0.67$, $p = 0.50$]; SP_happy vs. PO_happy: [$\beta = 0.07$, SE = 0.04, $t = 1.79$, $p = 0.07$]). Angry and happy stimuli were rated as more arousing when the emotion was conveyed by prosody only compared to when conveyed by semantics only (PO_angry vs. SO_angry: [$\beta =$ 0.88, SE = 0.04, $t =$ 22.79, $p < 0.001$]; PO_happy vs. SO_happy: [$\beta =$ 0.85, SE = 0.04, $t =$ 22.08, $p < 0.001$]). Furthermore, in female listeners, when the emotion was conveyed by semantics only, neither angry nor happy stimuli (spoken with neutral prosody) differed significantly in arousal from the neutral baseline (SO_angry vs. Baseline_neutral: [$\beta = 0.05$, SE = 0.04, $t = 1.34$, $p = 0.18$]; SO_happy vs. Baseline_neutral:[$\beta = 0.07$, SE = 0.03, $t = 2.19$, $p = 0.03$]). In male listeners, anger conveyed by semantics only did not differ from the neutral baseline ($\beta = 0.08$, SE = 0.04, $t = 2.09$, $p = 0.04$), whereas happiness conveyed by semantics only showed a significant increase in arousal compared to the neutral baseline ($\beta = 0.13$, SE = 0.04, $t = 3.27$, $p < 0.01$). No other significant sex differences were observed across conditions ($\beta = 0.02$, SE = 0.09, $t = 0.17$, $p = 0.86$).

As for valence ratings, angry stimuli were judged as less pleasant than neutral stimuli, whereas happy stimuli were

Table 4 Descriptive statistics and p-values for arousal, valence and SCRs

	Female listeners			Male listeners		
	Mean	SD	*p*	Mean	SD	*p*
Arousal						
Baseline_neutral (intercept)	1.78	0.87	**<0.001**	1.81	0.88	**<0.001**
SO_happy	1.93	0.97	0.03	2.03	0.86	**<0.01**
PO_happy	4.07	0.78	**<0.001**	4.03	0.76	**<0.001**
SP_happy	4.31	0.61	**<0.001**	4.33	0.72	**<0.001**
SO_angry	1.85	0.83	0.18	1.95	0.88	0.04
PO_angry	4.25	1.12	**<0.001**	4.40	0.75	**<0.001**
SP_angry	4.35	1.11	**<0.001**	4.43	0.74	**<0.001**
Valence						
Baseline_neutral (intercept)	2.86	0.50	**<0.001**	2.97	0.43	**<0.001**
SO_happy	3.24	0.67	**<0.01**	3.52	0.83	**<0.001**
PO_happy	3.91	0.79	**<0.001**	3.90	0.83	**<0.001**
SP_happy	4.55	0.69	**<0.001**	4.69	0.48	**<0.001**
SO_angry	2.38	0.75	**<0.001**	2.31	0.76	**<0.001**
PO_angry	1.95	1.02	**<0.001**	1.94	0.97	**<0.001**
SP_angry	1.63	1.11	**<0.001**	1.30	0.56	**<0.001**
SCRs						
Baseline_neutral (intercept)	0.44	0.49	**<0.001**	0.40	0.47	**<0.001**
SO_happy	0.55	0.59	0.03	0.41	0.53	0.92
PO_happy	0.53	0.65	0.12	0.31	0.36	0.25
SP_happy	0.50	0.49	0.11	0.40	0.57	0.90
SO_angry	0.54	0.67	0.02	0.37	0.38	0.96
PO_angry	0.84	0.72	**<0.001**	0.43	0.45	0.49
SP_angry	0.81	1.10	**<0.001**	0.43	0.47	0.44

P-values refer to the comparison between emotional cue conditions and the neutral baseline. P-values less than 0.017 were considered statistically significant. Significant effects are in bold judged as more pleasant than neutral ones. This effect was independent of the cue, i.e., whether the emotion was conveyed by the combination of semantics and prosody (SP_angry vs. Baseline_neutral: [β = -0.71, SE = 0.04, t = -17.80 $p < 0.001$]; SP_happy vs. Baseline_neutral: [$\beta = 0.46$, SE = 0.04, $t = 11.73$, $p < 0.001$]), by prosody only (PO_angry vs. Baseline_neutral: [β = -0.50, SE = 0.03, t = -14.09, $p < 0.001$]; PO_happy vs. Baseline_neutral: [$\beta = 0.31$, SE = 0.03, $t = 8.68$, $p < 0.001$]) or by semantics only (SO_angry vs. Baseline_neutral: [β = -0.23, SE = 0.04, t = -5.78, $p < 0.001$]; SO_happy vs. Baseline_neutral: [$\beta = 0.12$, SE = 0.04, $t = 3.04$, $p = 0.01$]). A comparison across cue conditions showed more extreme valence ratings (i.e., the most unpleasant for angry stimuli, the most pleasant for happy stimuli) when the emotion was conveyed by the combination of semantics and prosody compared to when the emotion was conveyed by prosody only (SP_angry vs. PO_angry: [β = -0.20, SE = 0.04, t = -5.17, $p < 0.001$]; SP_happy vs. PO_happy: [$\beta = 0.16$, SE = 0.04, $t = 3.95$, $p < 0.001$]). Furthermore, valence ratings were more extreme (i.e., more unpleasant for angry stimuli, more pleasant for happy stimuli) when the emotion was conveyed by prosody only than by semantics only (PO_angry vs. SO_angry: [β = -0.27, SE = 0.04, t = -6.85, $p < 0.001$]; PO_happy vs. SO_happy: [$\beta = 0.19$, SE = 0.04, $t = 4.74$, $p < 0.001$]). No significant differences were found between female and male listeners across conditions ($\beta = 0.04$, SE = 0.04, $t = 0.97$, $p = 0.33$).

SCR amplitudes

Figure 3 illustrates SCRs as a function of emotional cue and listener sex.

Larger skin conductance responses were found only for angry stimuli, but this effect was modulated by the sex of the listeners and the emotional cue. Specifically, for female listeners, SCR amplitudes increased for angry stimuli compared to neutral stimuli; this occurred both when anger was conveyed by the combination of semantics and prosody ($\beta = 0.21$, SE = 0.04, $t = 5.07$, $p < 0.001$) and when conveyed by prosody only ($\beta = 0.26$, SE = 0.04, $t = 6.17$, $p < 0.001$). No significant difference was found between these two cue conditions (β = -0.05, SE = 0.04, t = -1.27, $p = 0.20$). Angry semantics produced smaller SCR amplitudes than angry prosody (β = -0.16, SE = 0.04, t = -3.87, $p < 0.001$) and did not differ from neutral stimuli ($\beta = 0.10$, SE = 0.04, t = 2.25, $p = 0.02$). For male listeners, there was no difference between angry and neutral stimuli, independent of the emotional cue (all $p > 0.017$) Furthermore, no significant differences in SCR amplitudes were observed between happy stimuli and the neutral baseline, independent of the type of emotional cue or the sex of the listeners (all $p > 0.017$).

Correlation between behavioural ratings and SCRs

Figure 4 presents the correlation matrices for behavioural ratings and SCRs. The analysis also included arousal ratings of prosody from the validation task, as angry prosody was rated as more arousing than happy prosody. Correlation analyses were performed and plotted separately for female and male listeners, reflecting the observed interaction with listener sex.

The correlation analysis showed significant relationships between the variables. Among female listeners, a positive correlation was found between SCRs and the arousal scores of prosody from the validation task ($r_\tau = 0.14$, $p = 0.01$), as well as between SCRs and arousal scores from the main task (r_τ = 0.12, $p = 0.01$), indicating that listeners' SCRs increased in response to more arousing stimuli. Furthermore, SCRs were negatively correlated with valence (r_τ = -0.14, $p = 0.01$), meaning that listeners' SCRs increased when listening to unpleasant stimuli (which received lower scores on the valence scale). Additionally, the arousal scores of the prosody from the validation task were positively correlated with the arousal ratings from the main task (r_τ = 0.76, $p = 0.01$). This suggests that stimuli perceived as more arousing in terms of prosody were also rated as more arousing in the main task (when listeners were asked to rely on both semantics and prosody to rate the emotional dimensions of the stimuli).

Conversely, among male listeners, no correlation was found between SCRs and behavioural ratings. However, a positive correlation was found between the arousal of prosody and the rating of the arousal in the main task (r_τ = 0.82, $p = 0.01$), whereas a negative correlation was found between the arousal of prosody and valence ratings (r_τ = -0.15, $p = 0.01$).

Discussion

This study investigated how emotional semantics and/or prosody influence listeners' behavioural and skin conductance responses (SCRs). Behavioural results indicated that listeners primarily relied on prosody to assess arousal. In contrast, valence judgments were based on an integration of both semantics and prosody, with prosody exerting a stronger influence than semantics when the cues were presented independently. These effects were consistent across emotion categories and did not vary substantially by the sex of the listeners. In terms of physiological responses, increased SCRs were observed only in response to angry prosody, and this effect was specific to female listeners, regardless of the semantic content. The discussion is organised into two

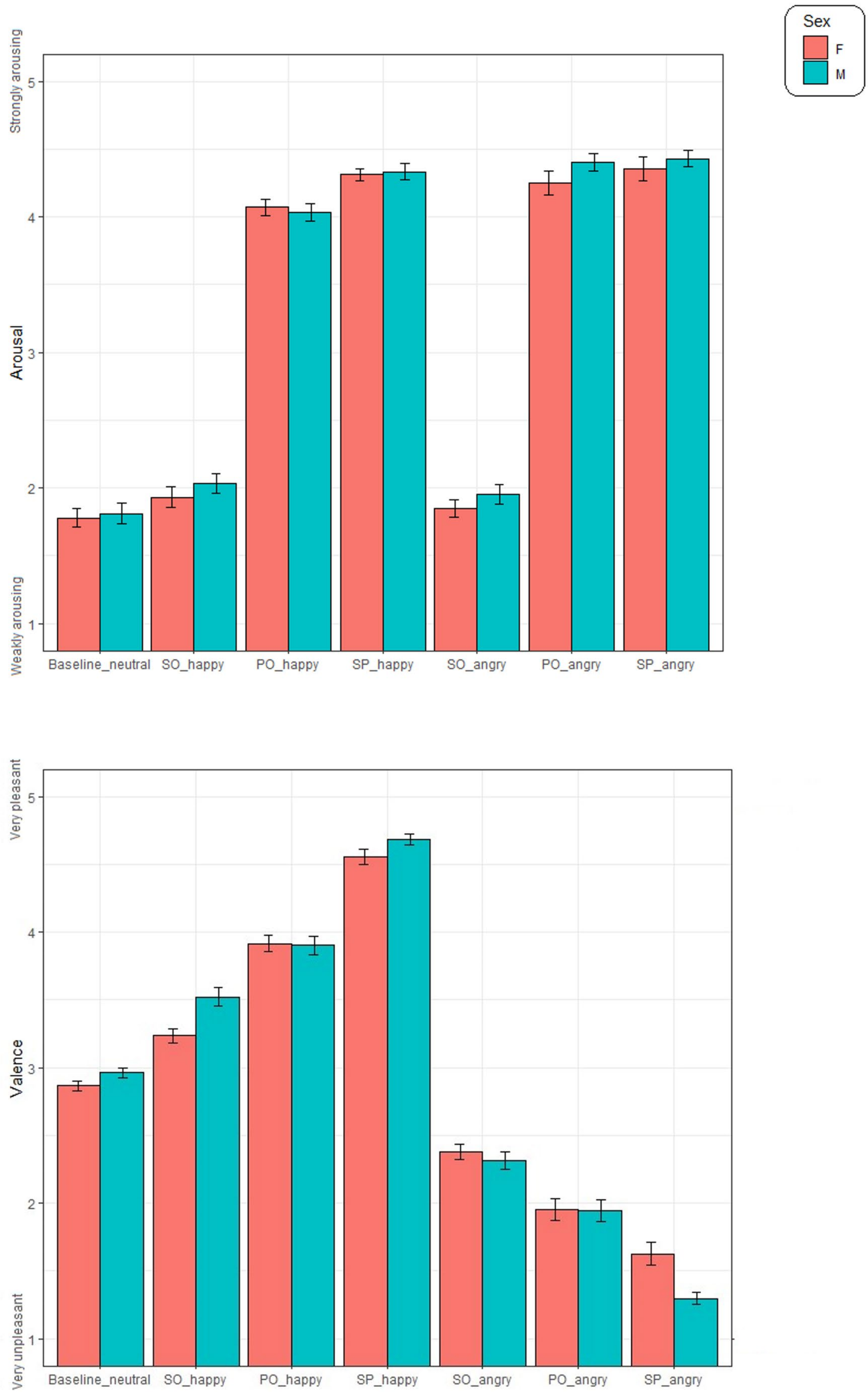
Sex
F
M
Strongly arousing
Weakly arousing
Arousal
1
2
3
4
5
Baseline_neutral
SO_happy
PO_happy
SP_happy
SO_angry
PO_angry
SP_angry
Very pleasant
Very unpleasant
Valence
1
2
3
4
5
Baseline_neutral
SO_happy
PO_happy
SP_happy
SO_angry
PO_angry
SP_angry

Fig. 2 Means and standard errors of arousal (top) and valence ratings (bottom) for the Emotional cue factor (seven levels: Baseline_neutral, SO_happy, PO_happy, SP_happy, SO_angry, PO_angry, SP_angry). Different colours illustrate results across listener sex (F = female listeners, M = male listeners)

sections: we first address the behavioural findings and then turn to the SCR results.

Behavioural responses

In the behavioural task, listeners were asked to rate the emotional arousal and valence of the stimuli, taking into account both semantics and prosody. We hypothesised that emotional ratings would be most extreme for stimuli that combined emotional semantics and prosody, followed by those conveying emotion through prosody alone. Interestingly, the results revealed distinct patterns depending on whether arousal or valence was being evaluated. Specifically, listeners predominantly relied on prosody to assess arousal, whereas both prosody and semantics contributed to valence ratings. These findings suggest that prosody and semantics play different roles depending on whether listeners are judging arousal or valence.

For arousal, listeners relied almost exclusively on prosody. They rated the utterances conveying anger and happiness through the combination of semantics and prosody and through prosody only as more arousing compared to utterances conveying anger and happiness through semantics only. No difference in arousal was found between anger and happiness conveyed by the combined cues and by prosody alone. This suggests that the addition of emotional semantics did not significantly increase perceived arousal when paired with emotional prosody. These findings are consistent with previous research suggesting that emotional prosody can override emotional semantics when evaluating stimuli with incongruent combinations of prosody and semantics (e.g., pleasant words or sentences spoken with unpleasant prosody; Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020). The prominent effect of prosody is further supported by the finding that utterances conveying anger through semantics alone (but spoken with neutral prosody) were rated as arousing as the neutral baseline by both female and male listeners. Despite being instructed to consider both semantic and prosodic cues, listeners may have relied more heavily on the neutral prosody to infer arousal, thereby overlooking the angry content conveyed through semantics. Interestingly, male participants rated happy semantic stimuli (spoken with neutral prosody) as significantly more arousing than the neutral baseline, whereas this difference was not statistically significant in female participants. This finding may suggest that men were more influenced by the positive verbal content of the utterances and did not rely solely on the neutral prosody to assess arousal. In contrast, women may have been more influenced by the neutral prosody of these stimuli, resulting in lower arousal ratings for the same semantic content.

For valence ratings, utterances conveying emotion through the combination of semantics and prosody received the most extreme scores. Specifically, anger conveyed through combined cues was rated as more unpleasant than anger conveyed through prosody or semantics alone. Similarly, happiness conveyed through both prosody and semantics was rated as more pleasant than when conveyed through prosody only. These heightened ratings for congruent emotional cues likely reflect a *congruency effect*—a phenomenon widely reported in previous research showing that emotional expressions are more effectively processed when prosodic and semantic cues are congruent and combined (e.g., Paulmann & Pell, 2011; Ben-David et al., 2016; Filippi et al., 2017a; Lin et al., 2020).

Furthermore, utterances conveying anger and happiness through prosody only resulted in more extreme valence ratings compared to those conveying anger and happiness through semantics only, consistent with prior research (e.g., Ben-David et al., 2016; Filippi et al., 2017a). However, unlike the results for arousal, utterances expressing anger and happiness through semantics alone were rated significantly different from neutral utterances by both female and male listeners. This suggests that while prosody may exert a stronger influence, semantic information is not ignored in valence judgments. Taken together, these findings support the view that listeners integrate both prosodic and semantic information when assessing emotional valence. In contrast, arousal ratings appear to be driven primarily by prosodic cues, with little influence from emotional semantics.

SCRs

We hypothesized that, given its older phylogenetic origins, emotional prosody would be more effective than emotional semantics in eliciting autonomic responses such as fight/flight/freeze reactions (Zimmermann et al., 2013; Filippi et al., 2017b; Filippi, 2020). Our findings partially supported this hypothesis. Specifically, the effect of emotional prosody on the autonomic nervous system, as reflected in SCRs, was modulated by the emotion category (anger) and by listeners' sex (female listeners). Stimuli conveying anger through prosody alone or through the combination of semantics and prosody elicited the largest SCRs, but this effect was observed only in female listeners. No significant difference in SCR amplitude was found between angry prosody alone and the combined semantics-prosody condition, indicating that adding emotional semantics did not further enhance autonomic responses. Additionally, stimuli conveying

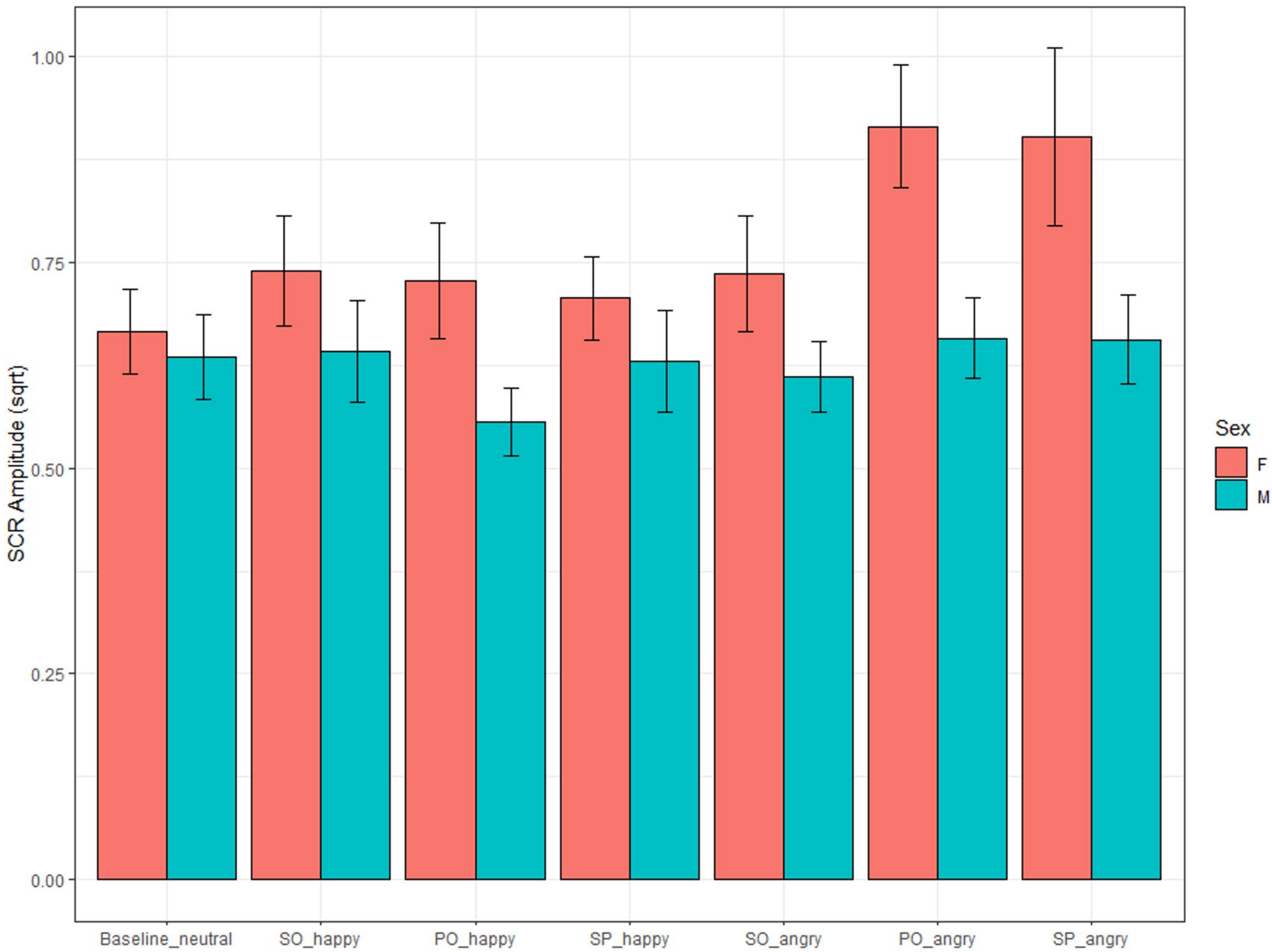

Fig. 3 Means and standard errors of SCR amplitude (sqrt) for emotional cue factor (seven levels: Baseline_neutral, SO_happy, PO_happy, SP_happy, SO_angry, PO_angry, SP_angry). Different colours illustrate results across listener sex (F = female listeners, M = male listeners)

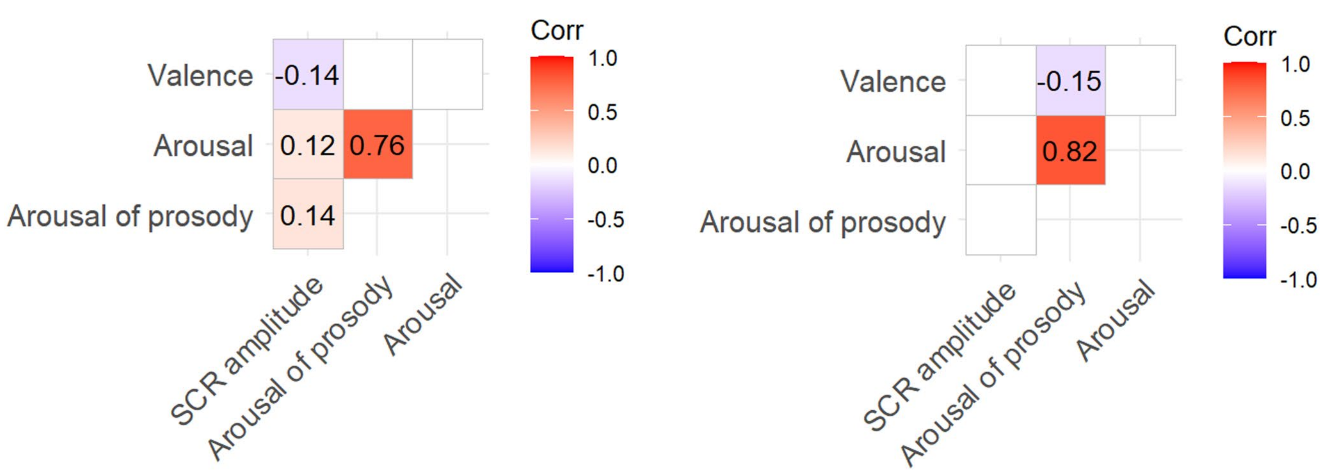

Fig. 4 Correlation matrices between behavioural ratings and SCRs for female listeners (F; left) and male listeners (M; right). Coloured cells indicate a significant correlation ($p < 0.01$), and values show Kendall's tau-b correlation coefficient. 'Arousal of prosody' refers to the arousal scores from the prosody validation task, while 'Arousal' and 'Valence' refer to the behavioural scores from the main experiment

happiness had no significant impact on SCRs, regardless of cue type or listeners' sex.

The observed sex difference in SCRs aligns with psychophysiological research using visual stimuli, which shows that women generally exhibit stronger autonomic responses than men to highly arousing unpleasant pictures, especially those depicting danger (Bradley et al., 2001; Chentsova-Dutton & Tsai, 2007; Williams & Gordon, 2007; Lithari et al., 2010; Gomez et al., 2016). For instance, Gomez et al. (2016) found that women displayed larger skin conductance after viewing pictures of unpleasant events, such as physical violence or mutilated bodies, whereas men showed stronger responses to highly arousing pleasant pictures, such as erotic pictures. Bradley et al. (2001) suggested that women may have a broader predisposition to exhibit heightened defensive activation to aversive cues compared to men, possibly due to biological factors (e.g., lower physical strength) and sociocultural influences (e.g., gender stereotypes). In contrast, while the erotic pictures used in Gomez et al.'s (2016) study were highly arousing, in our study, the lack of SCRs in men to happy prosody may be attributed to the arousal level induced by happy prosody not being high enough to trigger a physiological response. Indeed, as the validation tasks showed, happy prosody was rated as presenting an intermediate level of arousal. Alternatively, the lack of response may be attributed to differences in stimulus content, with erotic content in Gomez et al.'s study compared to happy content in the present study.

The effect of angry prosody, coupled with the absence of a significant effect for angry semantics on women's SCRs, suggests that women's autonomic responses are more sensitive to highly arousing unpleasant prosody than to equally arousing unpleasant semantics. This heightened sensitivity to prosody is further supported by the fact that combining angry semantics with angry prosody did not elicit stronger SCRs than prosody alone. As proposed by Aue et al. (2011), angry prosody may act as an evolutionarily salient acoustic signal that rapidly alerts listeners and prepares the body for action. While Aue et al. did not directly compare prosody and semantics, their findings support the notion that angry prosody alone can serve as a powerful trigger for physiological reactions.

Relationship between behavioural responses and SCRs

Although men and women provided mostly similar behavioural ratings of arousal and valence, men's autonomic responses, as indexed by SCRs, did not vary as a function of these emotional dimensions. Similarly, our correlation analysis revealed that, in women, SCRs were positively correlated with arousal ratings (from the behavioural task) and arousal ratings of prosody (from the validation task). However, these associations were not observed in men. These findings provide further evidence that women may be more responsive to both increased arousal and negative valence of the stimuli at the autonomic level, suggesting a heightened physiological sensitivity. In contrast, men's lack of response to these dimensions suggests that either they process emotional stimuli differently or that the level of emotional arousal induced by the stimuli in this study was insufficient to trigger a physiological response.

The partial dissociation observed between SCRs and behavioural responses supports the idea that the ANS activity may represent a distinct and partly independent pathway in emotional speech processing (Müller et al., 2011). The ANS plays a crucial role in preparing the body for adaptive responses to emotional stimuli by regulating physiological arousal, including changes in skin conductance, heart rate, and respiration (Critchley, 2009; Thayer & Lane, 2000). Unlike the cortical and subcortical brain areas typically emphasized in current models of emotional semantics and prosody processing (Schirmer & Kotz, 2006; Kotz & Paulmann, 2011), ANS responses might reflect more automatic, reflexive processes involved in the rapid detection of emotionally salient cues, particularly those related to threat or safety (Critchley & Garfinkel, 2017). This may suggest a parallel processing route that shapes emotional experience in ways not captured by higher-order cognitive processes reflected in cortical and subcortical activity. Moreover, the sex differences observed in SCRs to angry prosody highlight that autonomic responses can vary independently from conscious emotional evaluation. These findings underscore the importance of integrating ANS activity into comprehensive models of emotional speech processing and of conceptualizing emotion processing as a multi-level system where autonomic, subcortical, and cortical processes interact dynamically but can also diverge depending on individual and stimulus characteristics. Future research should explicitly investigate how these different levels communicate and influence one another to better understand the complexity of emotional speech processing.

Limitations, future directions and implications

In our study, we found no effect of emotional semantics on SCRs. This lack of effect may be attributed to the nature of the stimuli used in the experiment, which could have influenced the results. In the 'semantics only' condition, the contribution of semantics was isolated by using emotional sentences pronounced with neutral prosody. This choice was made to compare semantics and prosody on the same modality, i.e., the auditory one. However, it is possible that the arousal level of emotional semantics was mitigated by neutral prosody, thus impacting listeners' sensitivity to this emotional cue. As Pell et al. (2011) argued, the neutral tone of the utterance might have been perceived by listeners as incompatible with the interpretation of the emotional semantics, thereby introducing conflicting cues. Future studies should further investigate the impact of emotional semantics on SCRs by comparing emotional sentences presented in written form to those spoken with neutral prosody.

Additionally, we did not find an effect of happy prosody on SCRs. A limitation of our study is that utterances with happy prosody had lower levels of arousal compared to those with angry prosody (as shown by the validation tasks). This difference in arousal makes it difficult to isolate the specific contribution of valence in eliciting SCRs. Future studies should further clarify the role of valence in triggering SCRs by comparing happy and angry prosodies with similar levels of arousal. Furthermore, this disparity in arousal was not reflected in the overall acoustic proporties of the stimuli as no significant acoustic differences were found between angry and happy prosody. It might be possible that stimuli uttered with angry prosody contain salient acoustic changes in pitch and rhythm (such as high pitch excursions or modulations in the durational patterns) or specific intonation contours (in terms, e.g., of number and type of pitch accents) that might stimulate SCRs (Cao et al., 2014; Petrone et al., 2016). Future studies should further investigate the effects of phonetic and phonological aspects of angry prosody on the enhancement of SCRs.

Furthermore, the effect of angry prosody on SCRs was limited to female listeners. As the primary goal of our study was not to investigate sex differences in psychophysiological responses to emotional prosody and semantics, further research specifically addressing this concern is needed to validate our result. In this respect, a significant limitation of our experiment is that we used stimuli pronounced by one female speaker. Future studies should use stimuli pronounced by multiple female and male speakers to verify whether autonomic responses are modulated by idiosyncratic aspects of the speaker's vocal identity. One hypothesis may be that angry utterances spoken by male speakers may trigger higher SCRs than angry utterances spoken by female speakers. Indeed, men present lower f0 values than women. These lower f0 values have been associated with higher physical strength and dominance (Gussenhoven, 2015; Sell et al., 2010; Aung & Puts, 2020). Due to these acoustic characteristics, angry utterances spoken by male speakers might be perceived as more threatening, hence increasing listeners' SCRs.

While this study focused on anger and happiness expressions—primary emotions located at opposite ends of the valence spectrum (Ekman, 1992; Russell, 1980)—we acknowledge that this selection captures only a limited segment of the emotional range. Future research would benefit from examining a broader range of emotions, including those positioned along *affective continua*, such as anger versus calm or happiness versus sadness, to more precisely map how skin conductance varies with gradual shifts in arousal. Previous studies have shown that skin conductance scales with perceived arousal, with stronger SCRs elicited by highly arousing stimuli compared to low-arousing stimuli (Bradley & Lang, 2000; Lavezzo et al., 2024). However, it remains unclear whether similar arousal-driven patterns are observed when emotions, positioned along a *continuum*, are conveyed through the auditory channel via emotional semantics and/or prosody, and whether these responses are modulated by sex differences.

Similarly, more ambiguous emotions expressed through prosody and semantics may evoke distinct autonomic patterns compared to primary emotions. For instance, 'acid expressions', as defined by D'Errico and Poggi (2014), blend covert aggression, bitterness, and helplessness, and might elicit moderate increases in skin conductance compared to intense expressions of anger (e.g., hot anger), offering a valuable perspective on how emotional ambiguity modulates sympathetic reactivity. It would also be interesting to explore the impact of social emotions, such as pride and shame, conveyed through prosody and/or semantics on SCRs in both women and men. Some studies suggest that pride may be more strongly experienced by men, while shame is more prominent in women (e.g., Brody & Hall, 2008; but see Else-Quest et al., 2012). However, it remains unclear whether these differences are also reflected in the sympathetic nervous system, and, particularly when these emotions are conveyed through prosody and/or semantics.

Further research is also needed to determine whether our findings can be generalized across different cultures and languages. Previous studies suggest that speakers rely on emotional prosody and semantics differently depending on language and cultural context (Kitayama & Ishii, 2002; Ishii et al., 2003; Sauter et al., 2010; Paulmann & Uskul, 2014). For example, Japanese listeners, belonging to a high-context culture, tend to rely more on prosody, while American English speakers, belonging to a low-context culture, rely more on semantics to infer emotions (Kitayama & Ishii, 2002; Ishii et al., 2003). Moreover, research has shown that vocalizations, expressing basic emotions are universally recognized across diverse cultures (e.g., Western listeners and the Himba community), whereas some positive vocalizations (e.g., triumph, relief) show culture-specific recognition (Sauter et al., 2010). Cross-cultural comparisons of SCRs to emotional prosody and semantics could provide valuable insights into which aspects of auditory emotion processing are universal and which are influenced by cultural experience. We also note that our study focused on a single psychophysiological measure (SCR). Expanding the scope to include additional psychophysiological measures—such as heart-rate variability, pupil dilation, and facial electromyography—would provide a more comprehensive understanding of the psychophysiological mechanisms underlying the perception of emotional semantics and prosody.

Finally, our findings have both theoretical and practical implications. From a theoretical perspective, our study makes a significant contribution to understanding the autonomic responses involved in emotional speech processing. With these results, we take the first steps toward developing a peripheral physiology of emotional speech, laying the groundwork for future integration with existing models of emotional language processing, which have largely focused on brain and behavioural studies while often overlooking peripheral correlates due to limited evidence (Schirmer & Kotz, 2006; Kotz & Paulmann, 2011). Additionally, our study contributes to the ongoing debate on sex differences in emotional language processing (Schirmer & Kotz, 2006; Kotz & Paulmann, 2011; Thompson & Voyer, 2014) by providing psychophysiological evidence of sex differences in the processing of emotional prosody and semantics.

Our results also have practical implications, particularly in clinical settings. Indicators of sex differences in the processing of emotional prosody and semantics may serve as valuable diagnostic aids for conditions involving emotion regulation, such as internalising disorders like depression and anxiety, which are more prevalent in women compared to men (Albert, 2015; Salk et al., 2017). Indeed, individuals with major depressive disorder show a tendency to misinterpret emotional prosody, exhibiting a bias toward perceiving negative prosody and a blunting of positive prosody processing (Péron et al., 2011). Moreover, women with depression tend to engage more in rumination and may express more negative emotions in language compared to men (Nolen-Hoeksema, 2012). Together, these findings highlight the importance of integrating sex-specific patterns of emotional processing into diagnostic frameworks and therapeutic interventions to improve outcomes for individuals with emotional disorders.

Conclusion

This study investigated the impact of emotional semantics and prosody on behavioural and autonomic responses as reflected in SCRs. Our work provides evidence for the greater perceptual saliency of emotional prosody, which may be more effective than emotional semantics in transmitting emotional meanings in the auditory modality. We also reported psychophysiological data for the impact of angry prosody on female listeners, supporting the hypothesis of a sex differentiation in emotional speech processing at the autonomic level. Our findings align with the idea of emotional speech processing as a multistage process involving various processing levels, some of which partially driven by the autonomic nervous system and mediated by specific factors such as emotional dimensions of stimuli (e.g., arousal and valence) and the sex of listeners. Future studies are necessary to confirm whether and how listeners' behavioural and autonomic responses are affected by the sex and vocal identity of the speaker.

Acknowledgements This study was supported by a A*MIDEX grant to CP for the project 'COMEDIA', by a grant from the Agence Nationale de la Recherche (ANR) to CP for the project 'Representation and Planning of Prosody' (ANR-14-CE30-0005-01). This work, carried out within the Institute of Convergence ILCB, was also supported by grants from France 2030 (ANR-16-CONV-0002) and the Excellence Initiative of Aix-Marseille University (A*MIDEX). Furthermore, the study was supported by a mobility grant to FC as part of the PhD program at the University of Naples "L'Orientale". PF was supported by NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888. We thank Alain Ghio and Carine André for their help with data acquisition, as well as Erik Ringen and Robert Espesser for helpful comments on the statistical analyses.

Data availability The datasets generated during the current study are available on the Open Science Framework (OSF; https://osf.io/n5683/). The stimuli used in the experiment are available upon request from the author.

Declarations

Conflict of interest No potential conflict of interest was reported by the authors.

References

Albert, P. R. (2015). Why is depression more prevalent in women? *Journal of Psychiatry & Neuroscience*, *40*(4), 219–221. https://doi.org/10.1503/jpn.150205

Aue, T., Cuny, C., Sander, D., & Grandjean, D. (2011). Peripheral responses to attended and unattended angry prosody: A dichotic listening paradigm. *Psychophysiology*, *48*, 385–392. https://doi.org/10.1111/j.1469-8986.2010.01064.x

Aung, T., & Puts, D. (2020). Voice pitch: A window into the communication of social power. *Current Opinion in Psychology*, *33*(2), 154–161. https://doi.org/10.1016/j.copsyc.2019.07.028

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614. https://doi.org/10.1037/0022-3514.70.3.614

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological Sciences.* B275,473–481. https://doi.org/10.1098/rspb.2007.1460

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, V., & van Lieshout, P. H. (2016). Prosody and semantics are separate but not separable cues in the perception of emotional speech: Test for rating of emotions in speech. *Journal of Speech Language and Hearing Research*, *59*, 72–89. https://doi.org/10.1044/2015_JSLHR-H-14-0323

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, *86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Boersma, P. G. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 1572261550900588928

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer

Bradley, M. M., & Lang, P. J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology*, *37*(2), 204–215. https://doi.org/10.1111/1469-8986.3720204

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion*, *1*(3), 300. https://doi.org/10.1037/1528-3542.1.3.300

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20. https://doi.org/10.1111/j.1469-7998.2012.00920.x

Brody, L. R., & Hall, J. A. (2008). Gender and emotion in context. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 395–408). Guilford Press

Brown, S. (2017). A joint prosodic origin of Language and music. *Frontiers in Psychology*, *8*, 1894. https://doi.org/10.3389/fpsyg.2017.01894

Burriss, L., Powell, D. A., & White, J. (2007). Psychophysiological and subjective indices of emotion as a function of age and gender. *Cognition and Emotion*, *21*(1), 182–210. https://doi.org/10.1080/02699930600562235

Cacioppo, G., & Berntson, G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science*, *8*(5), 133–137. https://doi.org/10.1111/1467-8721.00031

Cao, H., Beňuš, Š., Gur, R. C., Verma, R., & Nenkova, A. (2014). Prosodic cues for emotion: analysis with discrete characterization of intonation. *Speech Prosody (Urbana, Ill.)*, *2014*, 130

Chentsova-Dutton, Y. E., & Tsai, J. L. (2007). Gender differences in emotional response among European Americans and Hmong Americans. *Cognition and Emotion*, *21*(1), 162–181. https://doi.org/10.1080/02699930600911333

Critchley, H. D. (2009). Psychophysiology of neural, cognitive and affective integration: fMRI and autonomic indicants. *International journal of psychophysiology*, *73*(2), 88-94. https://doi.org/10.1016/j.ijpsycho.2009.01.012

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current opinion in psychology*, *17*, 7-14. https://doi.org/10.1016/j.copsyc.2017.04.020

D'Errico, F., & Poggi, I. (2014). Acidity. The hidden face of conflictual and stressful situations. *Cognitive Computation*, *6*, 661–676. https://doi.org/10.1007/s12559-014-9280-1

Darwin, C. (1871). *The descent of man and selection in relation to sex*. John Murray

Dawson, M. E., Schell, A. M., & Filion, D. L. (2000). The electrodermal system. In: J.T. Cacioppo, L. G. Tassinary, G. G. Berntson (Eds.), *Handbook of psychophysiology*, Cambridge University Press, pp. 200–223.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P. (1999). Basic emotions. *Handbook of Cognition and Emotion*, *98*(45–60), 16.

Else-Quest, N. M., Higgins, A., Allison, S., & Morton, L. C. (2012). Gender differences in self-conscious emotions: A meta-analysis. *Psychological Bulletin*, *138*(5), 947–981. https://doi.org/10.1037/a0027930

Falk, D. (2004). Prelinguistic evolution in early hominins: Whence motherese? *Behavioral and Brain Sciences*, *27*(4), 491–503. https://doi.org/10.1017/S0140525X04000111

Filippi, P. (2020). Emotional voice intonation: A communication code at the origins of speech processing and word-meaning associations? *Journal of Nonverbal Behavior*, *44*(4), 395–417. https://doi.org/10.1007/s10919-020-00337-z

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., & De Boer, B. (2017a). More than words (and faces): Evidence for a Stroop effect of prosody in emotion word processing. *Cognition and Emotion*, *31*(5), 879–891. https://doi.org/10.1080/02699931.2016.1177489

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., & Güntürkün, O. (2017b). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: evidence for acoustic universals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Fischer, J. (2021). Primate vocal communication and the evolution of speech. *Current Directions in Psychological Science*, *30*(1), 55–60. https://doi.org/10.1177/0963721420979580

Fischer, J., & Price, T. (2017). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Gomez, P., & Danuser, B. (2010). Cardiovascular patterns associated with appetitive and defensive activation during affective picture viewing. *Psychophysiology*, *47*(3), 540–549. https://doi.org/10.1111/j.1469-8986.2009.00953.x

Gomez, P., Gunten, A., & Danuser, B. (2016). Autonomic nervous system reactivity within the valence–arousal affective space: Modulation by sex and age. *International Journal of Psychophysiology*, *109*, 51–62. https://doi.org/10.1016/j.ijpsycho.2016.10.002

Gussenhoven, C. (2015). *Foundations of intonation meaning: Anatomical and physiological factors*. Radboud University Nijmegen.

Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences*, *35*(5), 995–1005. https://doi.org/10.1016/S0191-8869(02)00313-6

Harris, C. L., Ayçíçeğí, A., & Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first Language than in a second Language. *Applied Psycholinguistics*, *24*, 561–579. https://doi.org/10.1017/S0142716403000286

Herrero, H., Tarrada, A., Haffen, E., Mignot, T., Sense, C., Schwan, R.,… and, & Hingray, C. (2020). Skin conductance response and emotional response in women with psychogenic non-epileptic

seizures. *Seizure*, *81*, 123–131. https://doi.org/10.1016/j.seizure.2020.07.028
Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods* (3rd ed.). Wiley.
Ishii, K., Reyes, J. A., & Kitayama, S. (2003). Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science*, *14*(1), 39–46. https://doi.org/10.1111/1467-9280.01416
Johnstone, T., & Scherer, K. R. (2000). Vocal communication of emotion. *Handbook of Emotions*, *2*, 220–235.
Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., & Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: A tutorial review and prospectus. *Biological Reviews*, *91*(1), 13–52. https://doi.org/10.1111/brv.12160
Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*, *16*(1), 29–59. https://doi.org/10.1080/0269993943000121
Kotz, S. A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, *1151*, 107–118. https://doi.org/10.1016/j.brainres.2007.03.015
Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Language and Linguistics Compass*, *5*(3), 108–125. https://doi.org/10.1111/j.1749-818X.2010.00267.x
Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010
Lavezzo, L., Gargano, A., Scilingo, E. P., & Nardelli, M. (2024). Inferring causality in emotions: A preliminary study on arousal perception and autonomic modulation. In 2024 IEEE International Symposium on Medical Measurements and Applications (MeMeA) (pp. 1–6). IEEE. https://doi.org/10.1109/MeMeA60663.2024.10596862
Levenson, R. W. (2003). Blood, sweat, and fears: The autonomic architecture of emotion. *Annals of the New York Academy of Sciences*, *1000*(1), 348–366. https://doi.org/10.1196/annals.1280.016
Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-cue and cross-modal Stroop effects. *Journal of Speech Language and Hearing Research*, *63*(3), 896–912. https://doi.org/10.1044/2020_JSLHR-19-00258
Lin, Y., Fan, X., Chen, Y., Zhang, H., Chen, F., Zhang, H., & Zhang, Y. (2022). Neurocognitive dynamics of prosodic salience over semantics during explicit and implicit processing of basic emotions in spoken words. *Brain Sciences*, *12*(12), 1706. https://doi.org/10.3390/brainsci12121706
Lithari, C., Frantzidis, C. A., Papadelis, C., Vivas, A. B., Klados, M. A., Kourtidou-Papadeli, C., & Bamidis, P. D. (2010). Are females more responsive to emotional stimuli? A neurophysiological study across arousal and Valence dimensions. *Brain Topography*, *23*, 27–40. https://doi.org/10.1007/s10548-009-0130-5
Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., & Eickhoff, S. B. (2011). Incongruence effects in crossmodal emotional integration. *Neuroimage*, *54*, 2257–2266. https://doi.org/10.1016/j.neuroimage.2010.10.047
Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261–289. https://doi.org/10.1007/BF02733986
Nolen-Hoeksema, S. (2012). Emotion regulation and psychopathology: The role of gender. *Annual Review of Clinical Psychology*, *8*, 161–187. https://doi.org/10.1146/annurev-clinpsy-032511-143109
Panksepp, J. (1982). Toward a general psychobiological theory of emotions. *Behavioral and Brain Sciences*, *5*(3), 407–422. https://doi.org/10.1017/S0140525X00012759
Paulmann, S., & Kotz, S. A. (2008). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo-and lexical-sentence context. *Brain and Language*, *105*(1), 59–69. https://doi.org/10.1016/j.bandl.2007.11.005
Paulmann, S., & Pell, M. D. (2011). Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion*, *35*, 192–201. https://doi.org/10.1007/s11031-011-9206-0
Paulmann, S., & Uskul, A. K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & Emotion*, *28*(2), 230–244. https://doi.org/10.1080/02699931.2013.812033
Paulmann, S., Jessen, S., & Kotz, S. A. (2009). Investigating the multimodal nature of human communication: Insights from erps. *Journal of Psychophysiology*, *23*(2), 63–76. https://doi.org/10.1027/0269-8803.23.2.63
Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, *25*(5), 834–853. https://doi.org/10.1080/02699931.2010.516915
Péron, J., Tamer, E., Grandjean, S., Leray, D., Travers, E., Drapier, D., Vérin, D., M., & Millet, B. (2011). Major depressive disorder skews the recognition of emotional prosody. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*(4), 987–996. https://doi.org/10.1016/j.pnpbp.2011.01.019
Petrone, C., Carbone, F., & Lavau-Champagne, M. (2016). May 31 - June 3, Effects of emotional prosody on skin conductance responses in French. Proceedings of Speech Prosody. Boston, MA, USA.
Roche, J. M., Peters, B. J., & Dale, R. (2015). Your tone says it all': The processing and interpretation of affective Language. *Speech Communication*, *66*, 47–64. https://doi.org/10.1016/j.specom.2014.07.004
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178. https://doi.org/10.1037/h0077714
Salk, R. H., Hyde, J. S., & Abramson, L. Y. (2017). Gender differences in depression in representative National samples: Meta-analyses of diagnoses and symptoms. *Psychological Bulletin*, *143*(8), 783–822. https://doi.org/10.1037/bul0000102
Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106
Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1–2), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5
Schirmer, A., & Kotz, S. A. (2006). Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, *10*(1), 24–30. https://doi.org/10.1016/j.tics.2005.11.009
Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*(2), 228–233. https://doi.org/10.1016/S0926-6410(02)00108-8
Schirmer, A., Striano, T., & Friederici, A. D. (2005). Sex differences in the preattentive processing of vocal emotional expressions. *Neuroreport*, *16*(6), 635–639. https://doi.org/10.1097/00001756-200504250-00024
Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., Von Rueden, C., & Gurven, M. (2010). Adaptations in humans for

assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1699), 3509–3518.

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of affective disorders*, *61*(3), 201-216. https://doi.org/10.1016/S0165-0327(00)00338-4

Thompson, A. E., & Voyer, D. (2014). Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cognition and Emotion*, *28*(7), 1164–1195. https://doi.org/10.1080/02699931.2013.875889

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org

Williams, L. M., & Gordon, E. (2007). Dynamic organization of the emotional brain: Responsivity, stability, and instability. *The Neuroscientist*, *13*(4), 349–370. https://doi.org/10.1177/10738584070130040801

Zimmermann, E., Leliveld, L. M. C., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: a comparative approach to mammalian voices, In E. Altenmuller, S. Schmidt, and E. Zimmermann (Eds.), *The Evolution of Emotional Communication*. Oxford University Press, pp. 116–132. https://doi.org/10.1093/acprof:oso/9780199583560.003.0008