It’s how you said it and what I heard: a comparison of motivational and emotional tone of voice

S. J. Haworth

A thesis submitted for the degree of Master of Science (by Dissertation) in Psychology

Department/School of Psychology

University of Essex

Date of submission for examination (September 2018)

Table of Contents

Table of Figures

Abstract

Previous research has viewed motivational and emotional vocal expressions as the same (e.g., Meyer & Turner, 2006; Fontaine & Scherer, 2013), but until now no direct comparison of these types of prosody has been available. Building on the new motivational prosody literature (e.g., Weinstein, Zougkou & Paulmann, 2014; 2018), this series of studies was the first to explore the differences and similarities between these forms of prosody. Initially, contextually valid sentences were intoned in angry, joyful, supportive, and controlling tones of voice by trained speakers, which were then acoustically analysed. Results revealed that each state was intoned with a different acoustic profile. Subsequently, exemplars were validated in a forced choice categorisation study and acoustics were extracted again. Results confirmed that each state was communicated with a different configuration of vocal cues, thus indicating that emotional and motivational states do not share the same prosodic profiles. In a final study, using an event-related potential (ERP) approach the time-course processing of these constructs was investigated. Findings suggest that emotional and motivational prosody share similar processing time-courses and neural resources. Weak evidence indicated possibly deviations in processing but were not strong enough to draw any conclusions. Taken together, the results of this investigation suggest that emotional and motivational prosody are likely distinct constructs. We conclude that these constructs differ on an encoding level and different vocal cues potentially lead to their effective recognition, but they are similar with respect to how they are processed in the brain. Implications, limitations and directions for future research are discussed. **253 words**

Keywords: Self-determination theory, motivational prosody, social prosody, emotional prosody, event-related potential.

It's how you said it and what I heard: a comparison of motivational and emotional tone of voice

Introduction

With the comprehensiveness in which tone of voice (also referred to as prosody) has been shown to augment vocal messages (e.g., Banse & Scherer, 1996; Murray & Arnott, 1993; Sobin & Alpert, 1999; Uskul, Paulmann & Weick, 2016) the informative power of vocal cues is difficult to dismiss. The addition of implicit information, through the manipulation of vocal cues (e.g., pitch, cadence, volume and speech rate; Banse & Scherer, 1996) has been shown to facilitate the communication of speaker affective states (Paulmann & Pell, 2010; Kraus, 2017) as well as attitudes and other social intentions (e.g., Cheang & Pell, 2008; Rigoulot, Fish & Pell, 2014; Weinstein, Zougkou & Paulmann, 2014, 2018).

Astonishingly, whilst it is accepted that prosodic communications have social as well as emotional functions (e.g., see Kreiman & Sidtis, 2013, for comprehensive discussion), compared to the extensive body of literature dedicated to emotional prosody (e.g., see Kotz & Paulmann, 2011; Paulmann, 2015 and Paulmann & Kotz, for cognitive and social neuroscience reviews), the prosodic communication of attitudes and other social intentions has been heavily neglected (e.g., see Mitchell & Ross, 2013, for review). Very recently, investigations into "motivational prosody" (e.g., Weinstein, Zougkou & Paulmann, 2014, 2018; Zougkou, Weinstein & Paulmann, 2017) have emerged. However, with arguably the largest deficit in the prosody literature being the lack of direct comparisons between types of prosody, only limited assertions regarding the distinctiveness of motivational tones of voice can be made. Although conceptually distinct (e.g., Batson, Shaw & Oleson, 1992), studies have

indicated that emotions and motivations are inseparably linked and that that emotions may play a fundamental role in the evocation of motivation (e.g., Scherer, 2004; Isen & Reeve, 2006; Meyer & Turner, 2006; Fontaine & Scherer, 2013; Vandercammen, Hofmans & Theuns, 2014). Conversely, some investigators conceptualise motivations as a component of an emotion (e.g., Fontaine, Scherer, Roesch & Ellsworth, 2007; Scherer, 1984, 1986).

Attitudinal and emotional prosody, which are also thought to be conceptually distinct (e.g., Mitchel & Ross, 2013; Wickens & Perry, 2015) have recently, on a decoding level, been shown to share a similar processing time-course and neural network (e.g., Wickens & Perry, 2015). Consequently, with no available direct comparison, at present any assertion made regarding the distinctiveness of motivational prosody is weakly grounded. By being the first investigation to directly compare motivational and emotional prosody with regard to encoding and decoding, the present investigation will provide a valuable insight into the differences and similarities between these constructs. As a result, this research will begin to rectify the deficit in the prosody literature, assist in the effective classification of prosody types and pave the way for more in-depth investigations of the finer aspects of prosodic communication.

Encoding (Production): Emotions

Extensive efforts have been made to identify which vocal cues are associated with the expression of discrete emotional states (e.g., van Bezooijen, 1984; Scherer, 1979; Scherer, Banse, Wallbott & Goldbeck, 1991; Banse and Scherer, 1996; Sobin & Alpert, 1999; Pell, Paulmann, Dara, Alasseri & Kotz, 2009; Castro & Lima, 2010; Lima & Castro, 2011; Paulmann & Uskul, 2014; Paulmann, Furnes, Boknes & Cozzolino. 2016; Also see Scherer, 1986 for review). Typically, informed by the

source filter model of speech production (Fant, 1960; Stevens, 2000), studies have predominantly described the same key attributes of speech (e.g., see Owren, & Bachorowski, 2007, for discussion): fundamental frequency (perceived as pitch) which indexes how high or low the voice sounds, intensity, also referred to as amplitude (i.e., how loud the voice is), rate of articulation (i.e., how quickly the utterance is conveyed), and voice quality (e.g., if the voice sounds crisp, breathy, grumbled, or harsh; for a detailed acoustic and physiological description of these parameters see, for example, Borden & Harris (1984)). However, some have argued that these key attributes are likely to be more reflective of non-specific physiological arousal than of discrete affective states (e.g., see Scherer 1979, 1986; Bachorowski & Owren, 2008, for discussions). Discussions on this topic tend to centre on the notion that these parameters account for a large proportion of the variance across all investigated emotions, whereas other acoustic measures (e.g., the Long term-average spectrum; LTAS, e.g., Pittam & Scherer, 1993) that are linked to distinct emotional states account for a great deal less of the variance (e.g., 10% in Banse & Scherer, 1996). Another interpretation could arguably be that these attributes constitute the acoustic foundation of all forms of prosody and that without them prosodic communication could not occur. In this view, their ability to account for the majority of the variance seems plausible but does not rule out the possibility that the expression of discrete emotional states is not contingent on these parameters. To be more specific, based on the empirical evidence in the literature (e.g., see Murray & Arnott, 1993; Banse & Scherer, 1996; Pittam & Scherer, 1993 for summaries) it seems fair to argue that distinct emotional states are expressed with a uniquely configured acoustic foundation (i.e., these key attributes are combined differently to create a unique acoustic profile). It is, however, important to note that some

evidence suggests the configuration of these vocal profiles is subject to the target language (e.g., Chinese, Arabic, Hindi, German and English; Pell, et al., 2009; Paulmann & Uskul, 2014). Much like facial displays of emotions which are argued to be bound by cultural display rules and expectations (e.g., Matsumoto, Consolacion, Yamanda, Suzuki, Franklin, Paul, Ray & Uchida, 2002), some evidence indicates that vocal displays may also be bound by cultural rules (e.g., Pell, Monetta, Paulmann & Kotz, 2009). Consequently, direct comparisons of acoustic profiles across languages may not be as insightful as previously thought. Nonetheless, accounting for language, findings are similar to the broader literature in the sense that distinct acoustic profiles have been reported for different emotions. For studies using English stimuli the findings can be summarised as follows (also see Table 1 for visual summary):

Disgust: Compared with neutral speech, disgust is reported to be characterised by a moderate increase in mean pitch, more pitch variability, a reduction in intensity, an increase in intensity range, with a slower rate of articulation and accompanied by a grumbled, chesty or nasal voice.

Fear: Consistently reported to be conveyed with a large increase in average pitch, a larger pitch range and an increase in intensity comparative to neutral speech. Fear is reported to be conveyed with an irregular voice quality, which some investigators have attributed to disturbances in respiratory patterns (e.g., Williams & Stevens, 1972). Rate of articulation findings have been inconsistent, being reported to either increase or decrease. This inconsistency may be a consequence of the form being studied, with decreases in speech rate likely attributed to milder forms of this emotion (e.g., anxiety or worry).

Happiness/joy: Unlike the broader literature, for studies related to English materials this emotion has been studied more in its subdued form (e.g., happiness) than its intense forms (e.g., elation or joy). Increases in average pitch and pitch range are reliably reported for happiness. Conversely, evidence suggests that joy is portrayed with a slight reduction in pitch, but with a large increase in pitch variability. In a similar vein, irregular, blaring and breathy voice qualities have been linked to this emotional category. Murray and Arnott (1993) report that joy contains large variations in stressed syllables (i.e., usually secondary syllables become stressed), thus it is likely that the blaring and irregular sounding voices are associated with the high arousal forms, such as joy. However, irrespective of its form, intensity and intensity range are reported to increase, as well as the presence of a slight to moderate increase in rate of articulation.

Sadness: Reliably reported to be expressed with an increase in mean pitch and pitch range, but a reduction in intensity, rate of articulation, and is accompanied by a resonant voice quality.

Anger: Generally conveyed with an increased fundamental frequency and pitch range, though there is some evidence of a reduction in average pitch. This emotion is also most commonly reported to be conveyed with an increase in both intensity and intensity variability, yet there is some evidence for a reduction in mean intensity (but not variability). With respect to voice quality, because of tense articulation (Murray & Arnott, 1993), anger is reported to be conveyed with a breathy voice quality and chesty tone. Findings for rate of articulation seem to be inconsistent across studies, with slight to moderate increases and decreases being reported. Given its high degree of recognisability, much of the disparity in acoustic cues for this emotion likely stems from the form in which it has been studied. Form

comparisons in the wider body of literature (e.g., cold vs. hot-anger; Banse & Scherer, 1996) show that the more subdued form of this emotion (cold-anger) is expressed with lesser increases in pitch, speech rate and intensity than its intense counterpart. In most cases, studies have not been transparent (i.e., explicitly stated) in which form of an emotion is being studied and as a result, renders the establishment of a robust acoustic profile more challenging.

***Table 1**: Summary of acoustic profiles for emotions.*

Parameter:	**Anger**	**Happiness**	**Sadness**	**Fear**	**Disgust**
Mean pitch	**<>**	**<>**	**>**	**>**	**>**
Pitch variability	**>**	**>**	**>**	**>**	**>**
Mean intensity	**<>**	**>**	**<**	**>**	**<**
Intensity variability	**>**	**>**	**<**	**>**	**>**
Speech rate	**=>**	**=>**	**<**	**<>**	**<**
Voice quality	Breathy Chesty	Breathy Blaring irregular	Resonant	Irregular	Chesty Nasal Grumbled

Note: < = Decrease; **>** = Increase; **<=/=>** = minor change; **<>** = change is reported in either direction; **<</>>** = large change; **=** = no change

Perhaps a more important debate surrounding the reliability of reported acoustic profiles across the literature is one of stimuli selection. In detail, some investigators have argued against the use of only “high quality” exemplars in acoustical analysis, labelling samples as unrepresentative and suggesting they are likely to inflate detection and recognition likelihood (Bachorowski & Owren, 2008). For instance, Banse and Scherer (1996) selected exemplars based on expert ratings (i.e., advanced students from a professional acting school) which resulted in 224 portrayals from 1344 being selected (17%). Similarly, Paulmann, et al. (2016), extracted 280 utterances from a possible 1155 (24%) using discriminate analysis. However, in cases such as the latter, subsets of recordings are used to decrease the presentation time of stimuli, thus is a consequence of the primary motivations of the

study. Though it is fair to acknowledge that in some cases screening processes are implemented as techniques for quantity reduction, rather than quality enhancement, they still yield a higher proportion of prototypical exemplars than would likely be expected in real life.

From a facial expression standpoint, cultural values and norms make prototypical visual displays of emotions scarce in real life (e.g., Matsumoto et al., 2002). If we apply this concept to the vocal communication of emotions, it seems unlikely that stimuli pools comprised of predominantly prototypical or “high quality” exemplars accurately reflect real emotional vocal expression. It also appears reasonable to assume that speaker variation and context differences can lead to less prototypical samples which are likely a more naturalistic reflection of how emotions are communicated through tone of voice. These less prototypical exemplars might also be more useful when testing for generalisable patterns in encoding (i.e., investigating which vocal cues speakers generally emphasise when expressing specific emotions).

On the other hand, considering that interpersonal relationships are heavily contingent on the accurate inference of affective states (e.g., Levenson & Ruef, 1992), with the exception of testing for generalisable cues in expression, the value of analysing poorly recognised exemplars is limited. Take for instance a speaker who intended to convey that they were angry through their tone of voice, but this was not picked up by the listener. Little to no social or interpersonal benefits would be conferred and it is likely that the expresser would need to rely on additional signals to ensure their message was received (e.g., facial cues, linguistics, contextual cues). Analysis of failed communications may provide some insight into which vocal cues speakers modulate in an attempt to convey a specific emotional state but does not

afford any understanding regarding how emotional states are effectively communicated through tone of voice. In fact, the accurate measurement of group differences (which in this case would be the different emotional states) requires stimuli that does not vary in the strength or intensity of emotion signalling properties (i.e., so that all stimuli are equally recognisable or obvious in expression characteristics; e.g., Matsumoto, 2002; 2007). If stimuli differ, comparisons are inextricably confounded by stimuli variation and as such the use of screened samples enables the effective assessment of group, rather than stimuli differences.

Evidently, this debate seems to lean toward the use of screened samples, as the vast majority of the literature has done. However, the very existence of the debate highlights the call for investigators to enhance the “realism” and representativeness of their samples. Arguably, by presenting acoustic data on the full set of recordings as well as the selected “high quality” stimuli, studies would be able to maintain the integrity of comparisons, have practically usable stimuli pools and also provide some insight into generalisable patterns in the vocal expression of emotions. Thus, the present study aims to provide the reader with data from both screened and unscreened samples. Hopefully, this will encourage future research addressing the questions of which vocal cues are used by speakers and which cues are used by listeners from a variety of angles.

Encoding (Production): Attitudes

Albeit limited to a handful of studies (e.g., Cheang & Pell, 2008; Rigoulot, Fish & Pell, 2014), evidence suggests that attitudes can be conveyed through distinctive tones of voice, which can be summarised as follows:

Sarcasm: Sarcastic utterances are expressed with overall reductions in mean pitch, respective to other attitudes and neutral speech. In general, sarcastic utterances are reported to be conveyed with an overall reduction in mean pitch, less pitch variation than sincerity, no change in average intensity or amplitude range and with more pronounced patterns of resonance and hoarseness. Evidence suggest that sarcastic speech rate varies depending on how it is linguistically transmitted (i.e., in short phrases such as “I suppose” or full sentences such as “It’s a respectful gesture”), with rate of articulation only reducing with regard to short phrases.

Sincerity: In contrast to insincere utterances (white lies), evidence suggests that sincerity is communicated with a reduction in average pitch, more pitch variability, a slight reduction in average intensity, more intensity range and a faster speech rate. Compared to neutral speech, sincerity is conveyed with a higher average pitch and with less hoarseness (measured by harmonics-to-noise ratio; HNR; Yumoto, Gould & Baer, 1982). Sincerity is reported to be expressed with less pronounced patterns of resonance than humour, sarcasm and neutral every day speech.

Humour: When compared to neutral speech, humour is communicated with an increase in average pitch, a large increase in pitch variability, with a higher mean amplitude and a reduction in amplitude variation. Interestingly, speech rate for this attitude appears to be moderated by speaker sex, with female speakers articulating humour slower than neutral speech and males increasing their rate or articulation. With humour shown to demonstrate the highest HNR values (compared to neutrality, sarcasm and sincerity; Cheang & Pell, 2008), this attitude appears to be communicated less hoarsely than other attitudes and neutral speech.

Noticeably, the differentiation of attitudes, just like for emotions seems to be highly contingent on similar cues. According to Pell (2006), the prosodic realisation of emotions and attitudes partially overlap in acoustic properties. This is not surprising considering that both affective states need to be communicated through the same available vocal attributes. Although the role of pitch modulation is important in both types of prosody, voice quality is reported to be crucial for the expression of emotions (e.g., Mitchell & Ross, 2013; Wickens & Perry, 2015), whereas attitudes are argued to be more contingent on pitch and rhythm modulation (e.g., Grichkovtsova, Morel & Lacheret, 2012). At present, with no studies prior to the present investigation directly comparing the acoustic profiles of emotions and attitudes, these reliances can only be inferred. The lack of any direct comparison has also opened the door for arguments of conflation of these constructs (e.g., Blanc & Dominey, 2003; Mozziconacci, 2001, also see Mitchell & Ross, 2013, for discussion). On a functional level, it has been posited that although vocal expressions of emotions may originate with the speaker's emotional state, they are not planned as emotional displays, but rather are social tools intended to influence the behaviour of others (e.g., Owen & Rendall, 1997; Owren, Rendall & Bachorowski, 2003; also see Russell, Bachorowski & Fernandez-Dols, 2003, for discussion).

Encoding (Production): Motivations

In this view, vocal expressions of emotions and motivations would be expected to not differ immensely. Very recently, investigations have begun to explore motivational tone of voice (e.g., Weinstein, Zougkou & Paulmann, 2014, 2018; Zougkou, Weinstein & Paulmann, 2017; Paulmann, Vrijders, Weinstein & Vansteenkiste, 2018) and has yielded findings that on the surface seem to contradict this point.

The interpersonal (i.e., the individual's drives for action) and intrapersonal (i.e., individuals influence the motivation of others to elicit a behavioural outcome) nature of motivation (Deci & Ryan, 2000), debatably makes it a pivotal component of the social communicative process via which social reciprocity occurs. Consequently, motivational prosody, much like attitudinal prosody, is an important medium for the transmission of social intentions and behavioural modification. Self-determination theory (SDT; Deci & Ryan, 1987; Ryan & Deci, 2000), upon which the motivational prosody literature is grounded, proposes that behaviour can be driven by two types of motivationally rich environments, controlling and autonomy supportive. Autonomy-supportive environments enhance feelings of choice and free expression (Niemiec & Ryan, 2009; Weinstein, Zougkou & Paulmann., 2014, 2018), inspiring others to action by promoting their well-being and self-endorsement of behaviours (e.g., Reeve, 2009; Weinstein & Ryan, 2010). Conversely, controlling environments stifle feelings of support, undermines self-expression and pays little regard to the well-being of the recipient (Soenens & Vansteenkiste, 2010). Controlling approaches attempt to energise others to act through pressure and coercion and have been linked to the effective provocation of immediate action (e.g., Bromberg-Martin, Matsumoto & Hikosaka, 2010), often without the message content being fully processed (e.g., Weinstein & Hodgins, 2009).

With respect to the two motivation qualities proposed by SDT, the following profiles have been reported (Weinstein, Zougkou & Paulmann. 2014; 2018):

Autonomy-supportive: In contrast to controlling tones, autonomy-supportive messages are expressed with a higher mean pitch, are quieter, have less loudness variation, are spoken slower and are conveyed in a softer tone of voice.

Controlling: Control is generally expressed with a more forceful voice quality, underpinned by more energy across frequency bands, is louder and exhibits greater fluctuations in intensity than autonomy-supportive prosody. Furthermore, controlling sentences are spoken more quickly and with a lower average pitch.

Comparisons to neutral non-motivationally laden speech (e.g., Zougkou, Weinstein & Paulmann, 2017) and evidence from small scale validation studies (e.g., Weinstein, Zougkou & Paulmann, 2014; 2018) suggests that these profiles differ from neutral (everyday speech). Similar to the emotional and attitudinal prosody literature, pitch, amplitude and speech rate differences were reported as distinguishing acoustic markers between motivational speech and the acoustic profile associated with neutral speech (e.g., Zougkou, Weinstein & Paulmann, 2017). Interestingly, however, so too was voice quality, which is thought to be pivotal in the expression of emotions (e.g., Mitchell & Ross, 2013; Wickens & Perry, 2015), but yet was often not looked at in previous prosody studies. Even though surface comparisons of emotional and motivational profiles indicate that they differ from one another (e.g., cold-anger and control differ on pitch direction and voice quality), this contrast is indirect and thus any difference could be a result of other factors (e.g., task, stimuli, encoders, or elicitation procedure). Consequently, to effectively establish the distinctiveness of motivational tones of voice, more comprehensive research that directly compares these types of prosody to the same neutral baseline, such as that offered by the present investigation is needed.

Decoding: Emotions

Given that effective intrapersonal communication is governed by the accurate transmission and inference of carefully selected information (Planalp, 1998), it is equally, if not more important that the intended message is perceived and

recognised by the target recipient. Take, for instance, the earlier example of a person trying to communicate anger to someone. If this was to remain unrecognised the speaker would remain in the same social situation that may have been the cause of this state.

Although not to the same level of accuracy as in situations where more information (e.g., through linguistic or visual cues) is available upon which to base their judgements (Paulmann & Pell, 2010), a strong body of literature demonstrates that prosody on its own is sufficient for humans to accurately discriminate and infer the emotional states of speakers (e.g., Kraus, 2017; Paulmann & Pell, 2010) and even arousal across species (e.g., terrestrial vertebrates; Flippi, et al., 2017). In fact, the vast majority of research that set out to identify the vocal configurations associated with the vocal expression of different emotions also incorporated some form of recognition or stimuli validation process, which in turn provides direct evidence in support of the claim that emotions can be recognised with better than chance accuracy through prosody alone (e.g., van Bezooijen, 1984; Scherer, 1979; Scherer, et al., 1991; Banse and Scherer, 1996; Sobin & Alpert, 1999; Pell, et al., 2009; Castro & Lima, 2010; Lima & Castro, 2011; Paulmann & Uskul, 2014; Paulmann, et al. 2016; Also see Scherer, 1986 for review). See Table 2 for a summary of recognition accuracy across 9 empirical studies.

Table 2: *Accuracy percentages for individual emotions across empirical decoding studies.*

Study	Fear	Disgust	Joy	Sadness	Anger
Van Bezooijen (1984) Chance level: 10%	58%	49%	72%	67%	74%
Scherer, et al. (1991) Chance level: 17%	52%	28%	59%	72%	68%
Banse & Scherer (1996) Chance level: 7%	36%	15%	*	52%	*
Sobin & Alpert (1999) Chance level 25%	59%	*	81%	74%	95%
Pell, et al. (2009) Chance level 14%	74%	68%	*	78%	79%
Castro & Lima (2010) Chance level 14%	60%	55%	*	83%	75%
Lima & Castro (2011) Chance level 14%	64%	40%	*	84%	59%
Paulmann & Uskul (2014) Chance level 14%	61%	60%	*	83%	86%
Paulmann, et al. (2016) Chance level 14%	62%	60%	*	71%	96%
Average:	58%	47%	71%	74%	79%

Note: * indicates that no directly comparable emotion was available.

Collectively these studies demonstrate that on the back of tone of voice alone, some emotions (e.g., anger and sadness) are more easily recognised than others (e.g. disgust and fear). Although consistent in the sense that none of these studies controlled for response bias, due to inconsistencies in empirical frameworks, comparison between studies should be done so with caution.

To begin with, instabilities in encoder quantity and criteria (e.g., lay, trained speakers, voice problems) - whilst arguably may be more reflective of a natural communicative environment - make comparisons of recognition accuracy increasingly difficult. Voice professionals (e.g., actors, singers) are reported to be better at modulating their voice (see Scherer, 1979 for paradigm discussion) and as such variation in lay and professional encoding may in fact be measuring recognition

in response to incomparable stimuli pools. In addition, with some studies using exemplars from as few as two encoders of the same sex (e.g., Castro & Lima, 2010; Lima & Castro, 2011) and others assessing recognition of stimuli from twelve encoders of both sexes (e.g., Banse & Scherer, 1996), disparities in recognition rates may be in part, a consequence of the different discrimination strategies used by judges in relation to the variation within target stimuli pools.

Similarly, a lack of uniformity in stimuli format (e.g., lexical sentences or pseudo-sentences) also raises concern over recognition rate findings. Predominantly, research in this field has embraced the use of pseudo-sentences or non-words (e.g., Banse & Scherer, 1996; Scherer et al., 1991; Pell, et al. 2009; Paulmann & Uskul, 2014), enabling the isolated investigation of affective recognition via prosody. Alternative approaches have included sentences that are free of strongly biasing words, but semantically valid (e.g., “The fence was painted brown”; van Bezooijen, 1984; Paulmann et al., 2016; Castro & Lima, 2010; Lima & Castro, 2011) and emotionally-laden sentences (e.g., “It’s hard to believe this is real. I can’t believe things like this happen”; Sobin & Alpert, 1999), both of which are arguably more naturalistic than pseudo-sentences, but produce more potential confounds. By reporting enhanced recognition accuracy for valid and meaningful sentences compared with pseudo-sentences, Castro & Lima (2010) illuminate how inconsistencies in stimuli format render comparison across studies problematic. To achieve a higher degree of naturalism, whilst maintaining consistency across comparisons, the present research will employ the same contextually relevant sentences, free of emotional and motivational biasing words (e.g., e.g., “Can you check this?” or “Tell me when you’re done”) across all categories of prosody.

Exacerbating the inability to confidently interpret recognition rates on face value, is disparity in stimuli language. Owing to the fact that emotional expressions and associated intensities are governed by cultural display rules (e.g., Matsumoto & Ekman, 1989; Matsumoto, et al., 2002), it is not especially surprising that recognition rates have been shown to differ across languages (e.g., Chinese, Arabic, Hindi, German and English; Pell, et al., 2009; Paulmann & Uskul, 2014). Of interest to the present study, are the investigations which utilised English stimuli. Interestingly, whilst studies using English stimuli largely differed from the broader literature in the positive emotion studied (i.e., studies using English stimuli switch from Joy to happiness), the pattern remains the same (i.e., some emotions were better recognised than others). See Table 3 for a comparison of studies in focus.

Table 3: *Empathetic accuracy across empirical decoding studies using English stimuli.*

Study	Fear	Disgust	Happiness	Sadness	Anger
Sobin & Alpert (1999) Chance level 25%	59%	*	*	74%	95%
Pell, et al. (2009) Chance level 14%	87%	76%	80%	90%	88%
Paulmann & Uskul (2014) Chance level 14%	64%	80%	48%	82%	91%
Paulmann, et al. (2016) Chance level 14%	62%	60%	40%	71%	96%
Average:	68%	72%	56%	79%	92%

Note: * indicates that no directly comparable emotion was available.

As discussed, comparison between different studies requires careful consideration of a large number of factors. Yet, collectively the literature indicates that different emotional states can be differentiated by listeners through prosody alone.

Decoding: Neurophysiology of emotions

Neurophysiological evidence portrays a similar picture to that of recognition studies, but also raises different questions. An extensive body of electroencephalography (EEG) research has investigated how the brain responds to incoming emotional messages. Collectively, the literature demonstrates that listeners can detect and differentiate between forms of emotional speech. However, there is less consensus on precisely how fast this occurs and what factors contribute to or mediate the process. Detection of emotional salience is predominantly reported to occur within 200ms of the utterance onset (e.g. Paulmann & Kotz, 2008; Paulmann, Ott & Kotz, 2011; Schirmer, Chen, Ching, Tan & Hong, 2013; Iredale, Rushby, McDonald, Dimoska-Di Marco & Swift, 2013). Emotional prosody processing is now accepted to be a complex multi-stage process (e.g., Schirmer & Kotz, 2006; Kotz & Paulmann, 2011), with the acoustic cues being extracted within the first 100ms after stimuli onset, which 100ms later undergo evaluation for salience and meaning. This is then followed by a later more cognitively dominated process of a more in-depth processing stage where the emotional meaning and details in the message are further evaluated (see Paulmann, 2015 and Paulmann & Kotz, In Press, for reviews). This in-depth processing stage has been linked to various directionally different (positive vs. negative) event-related potential (ERP) components, including the P300 (e.g., Bosantov & Kotchoubey, 2004; Iredale, et al., 2013), N400 (e.g., Wambacq & Jerger, 2004; Chang, Zhang, Zhang & Sun, 2018) late positive component (e.g., LPC; Schirmer, et al., 2013; Paulmann, Bleichner & Kotz, 2013; Stekelenburg & Vroomen, 2012) and other late negative potentials (e.g., Paulmann, Ott & Kotz, 2011). Interestingly, this component variability in the literature has been attributed to paradigm and stimuli differences (e.g., Paulmann 2015; Paulmann & Kotz, In Press)

and is accepted throughout to be reflective of enhanced processing of emotional attributes. More consistently, studies have reported varying P200 potentials when comparing emotional and neutral messages, however, there is also some evidence to suggest that emotions are also differentiated in this processing window (e.g., Paulmann, Bleichner & Kotz, 2013; Paulmann & Pell, 2010). For example, in Paulmann and Pell's (2010) study participants demonstrated potentials (N400-like) similar to those reported to reflect the final in-depth processing stage after being primed with 200ms or 400ms fragments of emotional sentences. In this study, it appeared that the short 200ms excerpts provided listeners with enough information to establish the emotional context and distinguish between emotions, thus implies that emotional categories can be inferred as quickly as 200ms after the start of a vocalisation. Very recently, a paper by Chang and colleagues (2018) assessed durational modulations on emotional speech processing in a tonal language. Participants were presented with semantically valid Chinese stimuli, that differed in duration (e.g., short: 0.5 -1 second; medium: 1.5-2 seconds; long: 2.5-3 seconds). Enhanced P200 responses were reported for anger and happiness compared with surprise, especially for shorter stimuli. Although the authors suggest that the demand of integrating information more quickly for shorter sentences would intuitively make enhanced P200 responses easier to find, the findings also nicely reinforce that differentiation of emotional categories can occur as quickly as 200ms onset.

Combined, though there is some disparity in exact ERP components and their precise functionality, the EEG literature and recognition rates imply that listeners rely on multiple acoustic cues when distinguishing and differentiating emotional prosodies (see Paulmann & Kotz, In Press, for a comprehensive discussion). However, the verdict on exactly what acoustic configuration (or profile) is needed to

distinguish a specific emotion from tone of voice is still out; a conclusion the present study will assist with reaching. Salience detection (P200 responses) has been reported to be sensitive to pitch (e.g., Pantev, Elbert, Ross, Eulitz & Terhardt, 1996), loudness (e.g., Picton, Woods, Baribeau-Braun & Healey,1977) and stimuli duration (e.g., Chang, et al., 2018) which behaviourally, could be operationalised as speech rate. Consequently, once again, these 'foundation' parameters serve as prime candidates for the basis of investigations into forms of prosodic communication on both, an encoding and decoding level.

Decoding: Attitudes

Although studies explicitly demonstrating that judges can accurately recognise attitudes through prosody alone are relatively scarce (e.g., Regel, Gunter & Friederici, 2010), but those that do, indicate that through the aforementioned 'foundation' parameters, listeners are able to discern different attitudes from the tone of voice of the speaker. However, most studies in this area have approached the phenomenon on a neurophysiological (e.g., ERP amplitudes) rather than cognitive-behavioural level (e.g., perceptual judgements). Although differences in these approaches need consideration, if we assume that in order for a judge to be able to consciously distinguish between different forms of prosody, their brain must have recognised the difference, these evidence bases become amalgamable. Viewed collectively, the evidence compellingly demonstrates that through only tone of voice, listeners can differentiate between different attitudes and social intentions. For instance, claims that social intentions such as sarcasm, irony and sincerity can be accurately inferred through prosody alone is convincingly reinforced by the respective neurophysiological findings (e.g., Regal, Gunter & Friederici, 2010; Rigoulot, Fish &

Pell, 2014; Matsui, Nakamura, Utsumi, Saski, Koike, Yoshida, Harada, Tanabe & Sadato, 2016).

Like for emotions, attitudes are assumed to undergo three stages of processing. Initially, vocal cues are extracted, after which utterances are evaluated for salience and later are subjected to more in-depth processing mechanisms. After a comprehensive review of neuroimaging and lesion data, Mitchell and Ross (2013) hypothesised that in the earlier basic decoding stages required for all forms of prosody, emotional and attitudinal prosody processing share the same neural resources, but differentiate in the later, more in-depth processing stage. Evidence for this later differentiation has been reported for ironic compared with literal sentences (e.g., Regel, Gunter & Friederici, 2010) and between insincere and sincere messages (e.g., Rigoulot, Fish & Pell, 2014), but when directly comparing emotional and attitudinal prosody processing Wickens and Perry (2015) failed to replicate this effect. Differentiation between ironic and literal sentences was reported to occur in a P600 component situated bilaterally in the posterior regions, whereas for sincerity the elicited P600 was reported in the right anterior region. Inconsistencies in effects linked to the other two processing stages also plague the attitudinal prosody literature. In the study by Regel and colleagues (2010) irony was reported to elicit negative effects at 250ms, whereas the effects reported by Wickens and Perry (2015) and Rigoulot, Fish and Pell (2014) were positive (P200) for sarcasm and sincerity. In a similar vein, while Rigoulot and colleagues (2014) reported expected N400 effects for sincerity, an absence of an N400 effect was reported by Regel, Gunter and Friederici (2010) in relation to sarcasm.

These diverse data patterns may be explained by research paradigms used across these studies. All three of these studies used very different research

paradigms. Regel, Gunter & Friederici (2010) presented target utterances following pragmatically different discourses. Rigoulot, Fish and Pell (2014) explored their target attitudes using a question and answer context in which participants heard responses to questions intoned in either a sincere or insincere tone of voice, whereas Wickens and Perry (2015) approached their investigation through an expectancy violation paradigm in which they cross-spliced sentences, combining neutral beginnings with sarcastic, neutral or angry endings. In addition, unlike in the other two studies, Wickens and Perry (2015) also manipulated task demands and whilst they report only minor differences between the tasks, their mere inclusion adds a further layer of complexity to the research paradigm, thus enhancing the likelihood that some of their findings may be a consequence of methodology. Also, worth considering is that in all cases the target attitude or social intention (i.e., sarcasm, sincerity, and irony) was one which is highly reliant on contextual information (e.g., “nice throw” could only be considered sarcastic if attached to an act of throwing in the immediate environment or in a shared experience). This notion is somewhat reinforced by the absence of the N400 component in the study by Regel and colleagues (2010), which they suggested indicated that irony presented with supportive contexts did not have difficulty integrating semantics. Therefore, it seems reasonable to suggest that as the contextual information presented through each of these approaches differs so enormously, their findings may be bi-products of how stimuli was presented, and thus emotional and attitudinal prosody processing may in fact differ as suggested by Mitchell and Ross (2013).

Decoding: Motivations

Drawing on a new domain of prosody research, namely motivational prosody, it seems as though emotions and social intentions may in fact be processed

differently in the brain, but these differences may not be quite as clear cut as Mitchell and Ross (2013) suggested. It should be acknowledged that a distinction between motivational and attitudinal prosody could be made, especially with respect to their reliance or interaction with contextual information. However, it is also the case that they both serve "social" rather than "emotional" functions (see Kreiman & Sidtis, 2013, for a comprehensive review and a discussion of this distinction). Moreover, the prosodic expression of attitudes is thought to be intentionally controlled (see Mitchell & Ross, 2013, for discussion), which is an argument that could also be made for motivations communicated via prosody.

It should come as no surprise that similar to emotions and attitudes, listeners have been shown to be able to correctly infer the intended motivational meaning of a message through prosody alone (e.g., Weinstein, Zougkou & Paulmann, 2014; 2018; Paulmann et al., 2018). To be more specific, the evidence demonstrates that listeners can recognise sentences that were intoned in a controlling tone of voice as more coercive and less supportive of choice than those expressed in an autonomy-supportive manner and neutral every day speech (e.g., Weinstein, Zougkou & Paulmann, 2014; 2018); an effect that was shown to persist across languages (e.g., Paulmann et al., 2018). Zougkou, Weinstein and Paulmann (2017) explored the time-course of these two qualities of motivation using prosody only and a combination of prosody and motivationally biasing words (e.g., "must"). For prosody only, they reported no evidence of motivational prosody processing in the very early window (80-170ms). Because N1 amplitudes have been reported to be susceptible to saliency evaluations (Liu et al., 2012), they attributed the absence of differentiating N1 amplitudes to the lack of obvious saliency cues in motivational tones of voice and highlight that speakers may vary pitch more when conveying motivations than

emotions, thus providing less consistent pitch cues. P200 amplitudes were found for the different motivational tones, with only controlling tones of voice eliciting significantly different amplitudes from neutral at this component, suggesting that controlling tones of voice are detected early and according to the authors are "tagged" as "important" or "motivationally relevant".

Interestingly, preferential processing of autonomy-supportive messages was observed when accompanied by motivationally biasing words, suggesting that autonomy-supportive messages are only "tagged" as important if they contain meaningful semantic content. As would be predicted by Mitchell and Ross's (2013) hypothesis, applied to prosodic expressions regulated by intention, controlling and autonomy-supportive messages were differentiated in the later higher-level processing stage (350-600ms), with more positive potentials elicited in response to controlling messages than for autonomy-supportive (which was reported to be trending from neutral). Overall, these findings indicate that motivational content of incoming stimuli is assessed for salience within 200ms of utterance onset and is tagged for preferential processing later in the time-course. Moreover, if conveyed with enough saliency cues, either through words or prosody, preferential processing is likely to occur. As was to be expected controlling tones of voice, which are acoustically harsher and are frequently used to elicit immediate behavioural outcomes (e.g., Weinstein & Hodgins, 2009; Bromberg-Martin, Matsumoto & Hikosaka, 2010), were harder to ignore and thus more easily lead to preferential processing. Perhaps most interesting, however, is the lack of differentiation in the N1 component. Whilst the N1 component was present in this study, the lack of significant differences between motivational qualities suggests that the cues extracted may have been similar and perhaps not specific to either motivational

quality. Because this was the first study to explore motivational prosody, further investigation is needed to assess ensure that motivational prosody does not elicit differential N1 responses and that the absence of this effect was not a methodological artefact. Yet taken as they are, these findings suggest that emotional and motivational prosody processing differs to a larger extent than only in the later components. This notion however, by directly investigating the time-course of motivational and emotional prosody in a shared methodological paradigm, is one that the present investigation aims to assess.

The present investigation

First and foremost, this set of studies set out to disentangle emotional and motivational prosody. To this aim, the present study will directly compare these forms of prosody through a consistent paradigm and with an identical neutral baseline. This comparison is novel in the sense that it will compare these constructs on both, an encoding and decoding level, thus providing a more comprehensive comparison. In doing so it is well situated to better assess whether these constructs are in fact distinct, inform the respective literatures, enhance the description of motivational acoustic profiles and assist with the accurate classification of prosody types.

As a subsequent concern, the present exploration aims to better inform 'real' social communication. Given that effective intrapersonal communication is governed by the accurate transmission and inference of carefully selected information (Planalp, 1998), listeners and vocalisers, alike need to be highly efficient in their transmission and recognition of prosodic information. It seems reasonable then, to propose that daily conversation imposes a large array of cognitive demands on both vocalisers and listeners. One such demand is that speakers need to be able to accurately

intone subtly different messages (e.g., anxiety and fear) and listeners are required to pick up on these delicate differences. To replicate this demand, states that are more subtly different and could easily be misconstrued as manifestations of each other will be explored in the present investigation. Cold-anger was selected as the negative emotion as this was considered to be more similar to controlling tones of voice than hot-anger, which is generally conveyed in a more intense manner. Similarly, joy was anticipated to share similar attributes to extreme manifestations of autonomy-supportive messages. Neutral prosody was included as a baseline measure, which for clarity will be categorised and referred to as a state through-out this research. Additionally, speakers need to be able to effectively convey their intended message upon different semantic structures, settings and even opposing semantic meanings (e.g., sarcasm; Cheang & Pell, 2008). Similar is true of listeners, for whom it is beneficial to recognise these tones of voice irrespective of the semantic content. To account for this within an empirical setting, we need to move away from isolating prosody through the use pseudo-sentences or semantically valid but contextually redundant stimuli. Instead, to break ground on informing 'real' social communications (i.e., how messages are conveyed through tone of voice in a real social setting), a better approach that will be taken by this research, may be semantically valid but not strongly category biasing sentences (e.g., "Tell me when you're ready", "Can you check this?").

A further demand conferred by real social interactions centres on idiosyncratic differences. In the real world, interactions between strangers are numerous and are laden with speaker variations in the way they choose to express a specific message. Irrespective of familiarity, both parties are required to effectively communicate and infer intended messages in order for the communication event to be successful. With

this in mind, listeners must be able to gleam intended meanings from messages across a wide range of speakers. With the influence of speaker idiosyncratic differences considered a key attribute of real-life communicative processes, an encoder pool that surpasses the majority of the previous literature (e.g., 12 speakers; Banse & Scherer, 1996) will be recruited for this research. Following Scherer (1979), voice professionals who are trained to modulate their voice will make up the encoder pool for these studies.

To ensure the best possible replication of real-life communicative demands, whilst maintaining empirical stability, the following studies will incorporate subtly different prosodic expressions from a larger than usual encoder pool. Utterances will be conveyed through semantically valid and contextually relevant sentences (i.e., sentences that could be freely used in a variety of contexts, such home, school, socialising and work without semantic modification). Sentences will be void of motivationally biasing words (e.g., “should” or “could”; Weinstein and Hodgins, 2009; Radel, Sarrazin and Pelletier, 2009) and emotionally-laden words (e.g., “wonderful” or “terrible”).

To ensure comparability with previous studies, three commonly shared indicators are of interest in the present investigation: pitch, loudness (or intensity) and speech rate. Following the motivational literature (e.g., Weinstein, Zougkou & Paulmann, 2014, 2018) the following bands in the spectrum are of interest as they have been linked to communicating voice quality features: 0-500Hz, 0-1000Hz, 500-100Hz, 1000-5000Hz, and 5000-8000Hz. This particular voice quality measure and spectrum bands were shown by Weinstein and colleagues (2014; 2018) to be a parameter upon which controlling, and autonomy-supportive tones of voice clearly differed, and as such were selected for inclusion in our acoustic analysis. With imbalances

between high and low frequency energy reported in the LTAS (i.e., less high frequency energy is present in the spectrum; e.g., Elowsson & Friberg, 2017), similar to previous studies that looked at voice quality (e.g., Banse & Scherer, 1996) overlapping energy bands were implemented, enabling the more precise measurement of potentially minor, but important differences in the lower frequency energy bands. Equally, to permit valid comparisons with past emotional, attitudinal and motivational prosody research the following ERP components are of interest: N1, P200, N400 and late positive or negative potentials.

Based on past prosody research and what is presently understood about the physiology of vocal production (e.g., harsh tones of voice are typically qualified by extreme tension in the vocal folds; see Mittal et al., 2013, for a review), predictions can be made for each of the states on the selected acoustic parameters. It is hypothesised that being distinct forms of prosodic communication, the acoustic profiles of all selected states will differ from each other and neutral speech on the chosen parameters, especially in terms of voice quality (i.e., voiced long-term average spectrum). In a similar vein, drawing on the neurophysiological emotional and motivational prosody literature, predictions regarding the expected components can be made. It is hypothesised that both anger and joy, but not autonomy-supportive or controlling messages will elicit varied N1 effects, thus emotions and motivations will differ in amplitudes at this component. Furthermore, we predict that control and anger will differentiate from neutral, but not from each other in the P2 window of interest. In the windows assessing the final processing stage (> 350 ms after stimulus onset) control and anger are expected to show different ERP amplitudes. Although all states will be included in analysis, anger and control will be selected as representatives of their respective prosody types upon which to ground

these hypotheses as both have been shown to elicit relatively strong amplitudes.

Acoustic and ERP predictions for all states are presented in Table 4 and 5.

Table 4: *Predicted state effects for selected acoustic parameters.*

Cold - Anger	Control	Joy	Autonomy - Supportive
Acoustic parameters:			
F0	>	<	>>	>
F0Ra	<	=	>	=
F0Var	<	=	>	=
IntRa	>	>	>	<
IntVar	>	>	>	<
Me**SR**	<	>	>	<
EB0.5K	<	>	>	<
EB1K	<	>	>	<
EB0.5-1K		>		<
EB1-5K		>		<
EB5-8K	>	=	=>	=
Note: < = Decrease; **>** = Increase; **<=/=>** = minor change; **<</>>** = large change; **=** = no change			
F0 = mean pitch (fundamental frequency), **F0**Ra = pitch range, **F0**Var = pitch variability, Me**Int** = mean intensity, **Int**Ra = intensity range, **Int**Var = intensity variability, Me**SR** = mean speech rate, Voice quality energy bands: **EB**0.5K = 0-500 Hz, **EB**1K = 0-1000 Hz, **EB**0.5-1K = 500-1000 Hz, **EB**1-5K = 1000-5000 Hz, **EB**5-8K = 5000-8000 Hz			

Table 5: *Predicted state effects for selected ERP components.*

Cold - Anger	Control	Joy	Autonomy - Supportive
ERP Component Time-Windows:			
N1	^	*x*	^	*x*
P2	^	*x*	^	^
late negativity (350-600ms)	^	^	^	^
Late long-lasting positivity (500-800ms)	*x*	^	*x*	^
Note: ^ = ERP amplitude elicited; ***x*** = ERP amplitude not elicited			

Study 1

The first study in our investigation set out to produce appropriately constructed stimuli for subsequent studies and assess whether there are generalisable differences in acoustic profiles across target states; in particular between emotional and motivational states. It has been argued that motivational qualities are likely to be less pronounced in an empirical setting than in real-life motivationally rich environments, where speakers would actually have to motivate another individual as opposed to imaging such a situation (Zougkou, Weinstein & Paulmann, 2017). Owing to the aims and planned methodologies of subsequent studies, in this study we used experienced speakers to intone the experimental stimuli. Being trained to modulate their voice, it was expected that the motivational and emotional components of the stimuli would be more pronounced and consequently be more in line with real-life occurrences of these expressions and better suited for the subsequent planned EEG study.

Semantic and syntactic content of the target sentences was identical across all states. Sentences to be intoned by encoders were carefully developed to have multi-context utility (i.e., all sentences were valid in more than one context with no modification needed) and were free of motivationally or emotionally rich words, thus rendering them less categorically biasing (e.g., “tell me when you’re done” and “why don’t we try tomorrow”).

Based on previous literature, both in motivational prosody (e.g., Weinstein, Zougkou & Paulmann, 2014, 2018; Paulmann et al., 2018) and emotional prosody (e.g., Sobin & Alpert, 1999; Pell, et al., 2009; Paulmann & Uskul, 2014) it was hypothesised that

acoustic profiles for all states will differ on pitch measures, and that emotions and motivations will generally be expressed in differing voice qualities.

Method

State Portrayals

Encoders

Fourteen native English speaker amateur actors (7 Female; Mean age = 20.07, Range = 18-24, SD = 1.94; Mean acting experience = 7.14, SD = 4.20) were recruited from the University of Essex theatre department and society to intone sentences to convey four distinct internal states (joy, cold anger, control, and autonomy-supportive) and neutral. They were paid £15 for their time.

Immersion Scenarios

As this study investigated five states (cold-anger, joy, control, autonomy-supportive and neutral) that were expected to potentially contain similar attributes (e.g., cold-anger being more restrained than hot anger may share similarities in acoustic cues with control, previously reported as harsher sounding; Weinstein et al., 2014; 2018), encoders were given scenarios to provide them with a better idea of the target state and to assist with immersion. Scenarios, depicting imaginary events in which the actors were described experiencing the given state in response to the event were developed for each state (except for neutral, where encoders were explicitly instructed to add no social intentions or emotions to their portrayals). State descriptions were laden with biasing words and phrases (e.g., “you felt a warm sensation in your tummy” for joy, “You wanted to grind your teeth in anger” for cold-anger, “They need to comply” for control, “They are under no obligation” for autonomy-supportive). Encoders were told to draw on personal experiences in which they felt similar to the actor in the scenario.

Non-biasing sentences

To effectively evaluate the contribution of prosody alone on the communication of internal states in the real world, 70 cross contextual sentences, free of strongly emotional and motivational biasing words were constructed. Each sentence was contextually relevant (i.e., could be used to convey meaningful information in the given context) in all intended contexts (e.g., workplace, education, home and in general social interactions with strangers, such as, customer service situations). 350 exemplars were recorded for each encoder (70 for each internal state), yielding a total of 4900 recorded sentences. 105 recordings were removed due to recording errors, leaving a final sample of 4795 exemplars. Recording errors included, artefacts in the recordings (e.g., crackles, clicks), poor clarity of speech (e.g., muffled or overly breathy), and the inclusion of additional or modified words (see Appendix 1 for full sentence list).

Recording procedure

All exemplars were recorded in a single session that began with encoders intoning neutral utterances, followed by the other internal states in a randomised order. For neutral recordings encoders were asked to read the sentences as they saw them “without any emotions or feelings”. Between conditions (including neutral), encoders were given a short break in which they were asked to discuss any aspect of their course they desired with the researcher. This discussion served as a distractor task, reducing state crossover. Prior to recording the affective and motivational states, encoders were presented with scenarios to assist with immersion in the target internal state. Encoders were provided with as much time as they needed to become immersed (1-2 minutes on average). Once immersed,

sentences were presented one at a time, allowing for repetition if the encoder was unhappy with the quality of their portrayal, or if an error was made. Sentences for each condition were presented in a fixed random sequence that was the same for all encoders.

Utterances were captured using a high-quality microphone (44.1 kHz, 16-bit, stereo) on Audacity software (ver. 2.2.1). Encoders were instructed to maintain a similar distance from the microphone for all conditions, which was monitored by the experimenter. Recordings were saved as sound files (wav) and Praat (Boersma & Weenink, 2018) was used to insert utterance boundaries.

Acoustic analysis

Prior to acoustic analysis, because of unavoidable differences in speaker intensity (i.e., some speakers were much quieter across all conditions), all audio files were normalised with mean amplitude set at a constant value, therefore only amplitude range will be looked at. Subsequently, to describe the acoustic typology of target states, the pitch (mean and range), intensity (range), speech rate (by syllable) and voice quality (voiced long-term average spectrum) were measured using customised scripts in Praat. Acoustics were extracted from audio files in their recorded format (.wav) with pitch floor and ceiling set differently for male and female encoders (75-450Hz for men, and 125-650Hz for women). For voice quality, proportions of energy in the 0-500Hz, 0-1000Hz, 500-1000Hz, 1000-5000Hz and 5000-8000Hz bands was measured.

Statistical analysis

Acoustics were analysed with separate mixed models, with by-encoder and by-utterance random intercepts corresponding to state. The fixed effect was the

target state and the outcome variables were the target acoustic parameters. Post-hoc contrasts ($p<.05$) used a modified Bonferroni correction for multiple comparisons (Keppel, 1991; $aplha_{corrected}$ = ($alpha_{base}$ x $df_{test\ condition}$) / number of comparisons). Elaborations of significant effects for state included 10 contrasts, resulting in a corrected p =.02.

As recording errors were not evenly spread across speaker sex, it was important to account for the extensively demonstrated sex differences in relation to pitch measures (e.g., Banse & Scherer, 1996; Cheang & Pell, 2008; Pell et al., 2009; Mittal, Erath & Pleniak, 2013; Monson, Hunter, Lotto & Story, 2014; Weinstein et al., 2014, 2018). As such pitch measures (mean and range) were normalised to reflect proportional changes for each speaker compared with their highly stable "resting frequency" (Menn & Boyce, 1982). Following Pell and colleagues (2009) normalised measures were calculated as follows: $F0Mean_{Norm}$ = ($F0mean_{Observed}$ – Resting frequency) / Resting frequency and $F0Range_{Norm}$ = (($F0max_{Observed}$ – Resting frequency) / Resting frequency) - (($F0min_{Observed}$ – Resting frequency) / Resting frequency). In their normalised formats, scores of 1 indicate that for a given utterance the speaker's mean pitch and/or expressive range was twice that of their resting frequency.

Results

Parameter averages for all exemplars across states are summarised in Table 6.

Table 6: *Mean (and SD) of acoustic parameters for all exemplars across states.*

Parameter	Anger	Control	Joy	Autonomy	Neutral
Pitch (normalised):					
Mean (F0)	0.42 (0.27)	0.45 (0.21)	0.85 (0.39)	0.58 (0.31)	0.30 (0.19)
Range	1.30 (1.04)	1.15 (0.83)	1.52 (0.78)	1.28 (0.84)	1.09 (1.03)
Amplitude (dB):					
Range	37.09 (9.04)	36.30 (8.70)	34.76 (8.32)	34.72 (8.04)	34.45 (7.36)
Speech rate	0.21 (0.05)	0.19 (0.05)	0.17 (0.04)	0.19 (0.04)	0.19 (0.04)
Energy Bands (Hz):					
0-500	39.67 (2.65)	39.73 (2.70)	39.21 (2.83)	40.08 (2.19)	40.29 (2.58)
0-1000	38.62 (1.98)	38.79 (1.92)	38.48 (1.93)	38.84 (1.69)	39.13 (1.68)
500-1000	35.37 (4.40)	35.91 (3.94)	35.78 (4.02)	35.30 (4.24)	35.80 (3.62)
1000-5000	24.95 (4.17)	25.35 (3.77)	25.58 (3.67)	24.36 (3.60)	24.12 (3.93)
5000-8000	11.27 (4.99)	11.79 (5.32)	12.64 (5.66)	12.02 (4.97)	10.86 (5.78)

Pitch

Mean pitch (F0)

A significant effect of *state* was found, $F(4, 4790) = 813.46$, $p < .001$. All states were communicated with significantly different average pitches ($p < .003$, $t > 2.965$ in all instances). Joy showed the largest proportional increase in average pitch (m = .85, *SD* = .39). The effects are illustrated in Fig. 1.

Pitch range

A significant effect of *state* was found, $F(4, 4790) = 813.46$, $p < .001$. With the exception of contrasts between autonomy and anger ($b = .01$, $t(4790) = 0.26$, $p = .794$) and control and neutral ($b = .06$, $t(4790) = 1.85$, $p = .064$), all other pairwise comparisons revealed significant differences in pitch range (all $p < .001$, $t > 3.625$).

Joy was expressed with significantly more variability in pitch than all other states (m = 1.52 *SD* = .78). Findings are illustrated in Fig. 1.

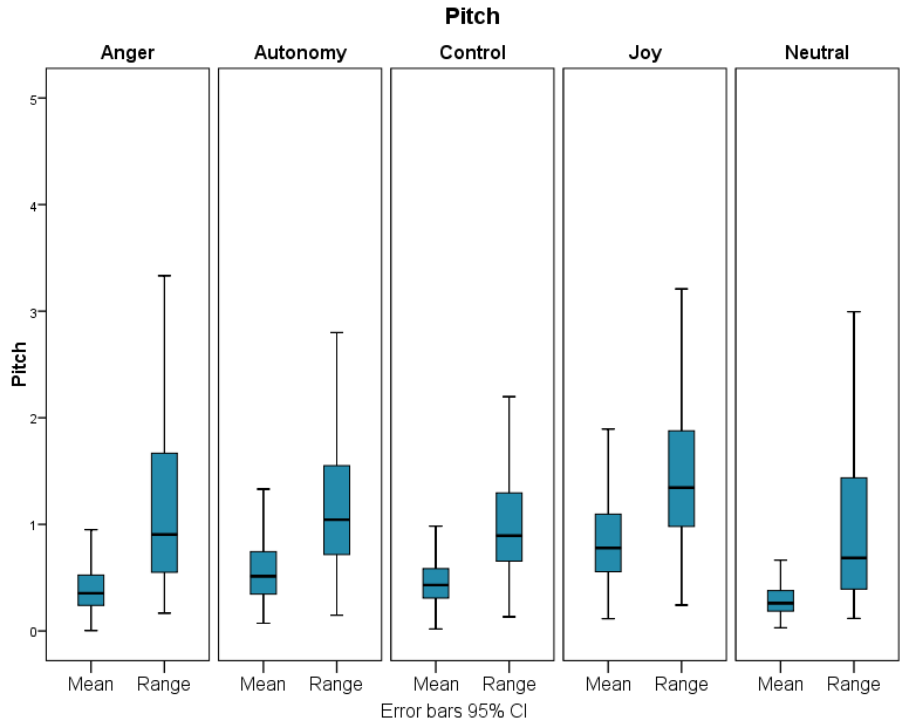

Figure 1: *Standardised mean pitch and range of all exemplars across investigated states.*
Note: *Measures were normalised in reference to induvial speaker resting frequencies.*

Amplitude

Amplitude range

With regard to amplitude range, a significant effect of *state* was found, $F(4, 4790) = 27.814$, $p < .001$. Angry portrayals demonstrated the largest amplitude range (m = 37.09, SD = 9.04) whereas neutral speech was found to contain the lowest amplitude range (m = 34.45, SD = 7.36). Contrasts demonstrated that differences in range for neutral, joy and autonomy were not significant (all $p > .412$, $t < .820$). Contrasts between other states were all significant ($p < .01$, $t > 2.481$ in all cases).

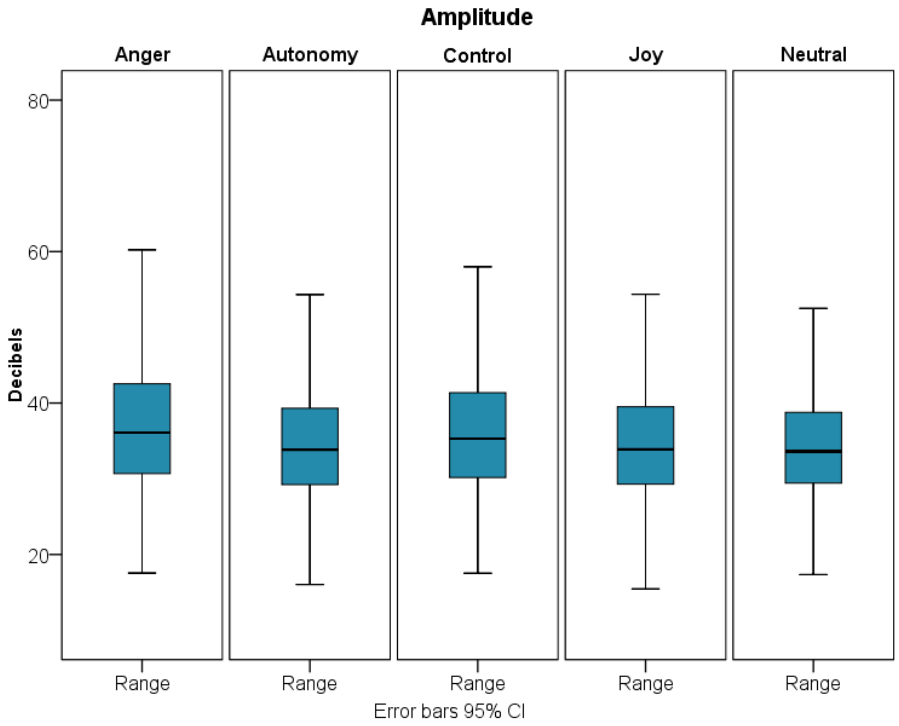

Figure 2: *Amplitude range for each of all exemplars across investigated states.*

Speech rate

Rate or articulation was shown to differ significantly as a function of *state, F*(4, 4790) = 186.100, *p* <.001. Joy was conveyed the fastest (m = .17, *SD* = .03) and anger had the slowest rate of articulation, taking 0.21 (*SD* = .05) seconds per syllable. Excluding neutral compared with autonomy (*b* = -.003, *t*(4790) = -1.95, *p* < .051) and control (*b* = .002, *t*(4790) = 1.79, *p* = .073), differences in rate of articulation were significantly different across states (*p* <.001, *t* > 3.742 in all instances). Duration per syllable across states is illustrated in Fig. 3.

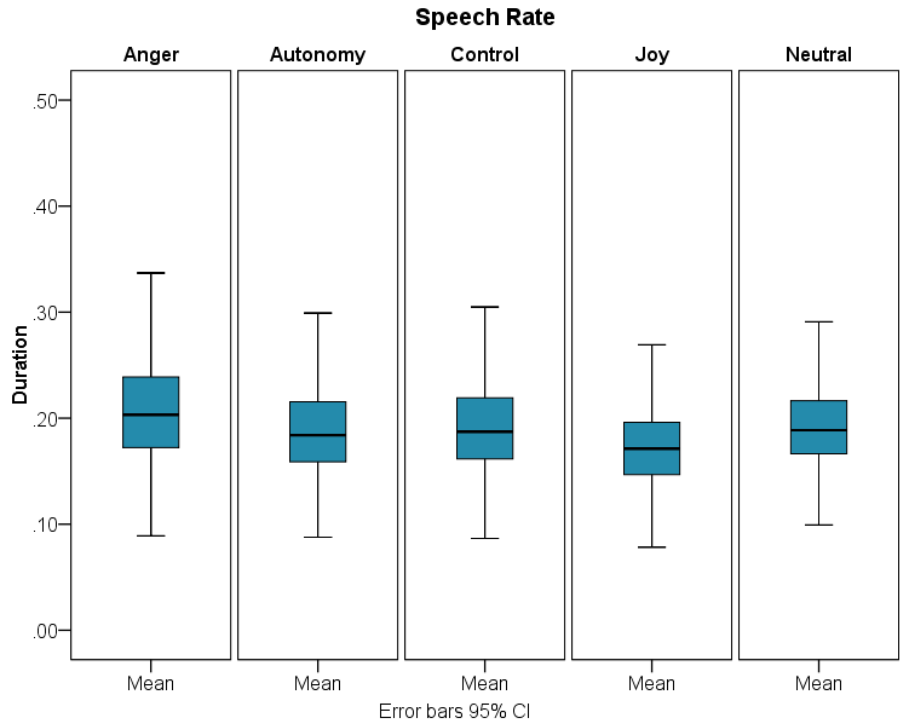

Figure 3: *Mean speech rate of all exemplars for each of the five investigated states.*
Note: *Higher values indicate a slower speech rate (duration per syllable).*

Voice quality

Significant effects of *state* were found across all measured bands of the long-term average spectrum (0-500Hz, $F(4, 4790) = 38.702$, $p < .001$; 0-1000Hz, $F(4, 4790) = 19.540$, $p < .001$; 500-1000Hz, $F(4, 4790) = 7.543$, $p < .001$; 1000-5000Hz, $F(4, 4790) = 53.251$, $p < .001$; 5000-8000Hz, $F(4, 4790) = 32.433$, $p < .001$). For clarity, pairwise contrast will be presented by state across all bands.

Anger – Autonomy

Encoders expressed anger with significantly more energy in the lower frequency bands (0-500Hz and 0-1000Hz) than autonomy ($b = -.414$, $t(4790) = -4.395$, $p < .001$ and $b = -.224$, $t(4790) = -2.874$, $p < .004$, respectively). Anger was spoken with higher energy in the 1000-5000Hz band ($b = .059$, $t(4790) = 4.854$, $p < .001$), but less in the 5000-8000Hz band ($b = -.733$, $t(4790) = -4.295$, $p < .001$).

Anger – Control

Expression of anger contained less energy than control at bands ranging from 500-1000Hz (*b* = -.527, *t*(4790) = -3.763, *p* < .001), 1000-5000Hz (*b* = -.410, *t*(4790) = -3.368, *p* < .001) to 5000-8000Hz (*b* = -.524, *t*(4790) = -3.079, *p* < .002).

Anger – Joy

Anger was communicated containing more energy at 0-500Hz (*b* = .458, *t*(4790) = 4.862, *p* < .001) and less energy in the 500-1000Hz (*b* = -.401, *t*(4790) = -2.856, *p* < .004), 1000-5000 (*b* = -.636, *t*(4790) = -5.211, *p* < .001) and 5000-8000Hz (*b* = -1.369, *t*(4790) = -8.019, *p* < .001) components of the spectrum.

Anger – Neutral

Angry utterances contained less energy in frequencies up to 1000Hz (0-500Hz, *b* = -.613, *t*(4790) = -6.516, *p* < .001; 0-1000Hz, *b* = -.499, *t*(4790) = -6.415, *p* < .001; 500-1000Hz, *b* = -.413, *t*(4790) = -2.953, *p* < .003) and more higher frequency energy than neutral (1000-5000Hz, *b* = .822, *t*(4790) = 6.747, *p* < .001; 5000-8000Hz, *b* = .413, *t*(4790) = 2.425, *p* < .015)

Autonomy – Control

Autonomy was expressed with more 0-500Hz energy *(*0-500Hz; *b* = .358, *t*(4790) = 3.819, *p* < .001) and significantly less energy at 500-1000Hz (*b* = -.603, *t*(4790) = -4.322, *p* < .001) and 1000-5000Hz (*b* = .-1.002, *t*(4790) = -8.255, *p* < .001).

Autonomy – Joy

Supportive utterances were expressed with significantly more energy in the 0-500Hz (*b* = .872, *t*(4790) = 9.283, *p* < .001) and 0-1000Hz (*b* = .366, *t*(4790) = 4.714, *p* < .001) bands. Versus joy, less energy was yielded by autonomy-supportive sentences in the 500-1000Hz (*b* = -.477, *t*(4790) = -3.411, *p* < .001), 1000-5000Hz (*b*

= -1.228, *t*(4790) = -10.093, *p* < .001) and 5000-8000Hz (*b* = -.636, *t*(4790) = -3.736, *p* < .001) bands

Autonomy – Neutral

Autonomy was conveyed with less energy than neutral between 0 and 1000Hz (*b* = -.275, *t*(4790) = -3.547, *p* < .001), mostly between 500-1000Hz (*b* = -.490, *t*(4790) = -3.509, *p* < .001). Autonomy also yielded more high frequency energy at 5000-8000Hz (*b* = 1.146, *t*(4790) = 6.795, *p* < .001).

Control – Joy

Versus joy, controlling messages were communicated with more energy in the 0-500Hz (*b* = .514, *t*(4790) = 5.482, *p* < .001) and 0-1000 (*b* = .307, *t*(4790) = 3.958, *p* < .001) bands. Control was expressed with less 5000-8000Hz energy than joy (*b* = -.844, *t*(4790) = -4.790, *p* < .001).

Control – Neutral

Control was conveyed with less 0-500Hz (*b* = -.557, *t*(4790) = -5.949, *p* < .001) and 0-1000Hz (*b* = -.334, *t*(4790) = -4.319, *p* < .001) energy and more energy in the 1000-5000Hz (*b* = 1.232, *t*(4790) = 10.161, *p* < .001) and 5000-8000Hz (*b* = .938, *t*(4790) = 5.529, *p* < .001) bands.

Joy – Neutral

Joyful exemplars contained less energy at 0-500Hz (*b* = -1.071, *t*(4790) = -11.416, *p* < .001) and 0-1000Hz (*b* = -.641, *t*(4790) = -8.266, *p* < .001) proportions of the spectrum. In the 1000-5000Hz (*b* = 1.457, *t*(4790) = 11.997, *p* < .001) and 5000-8000Hz (*b* = 1.782, *t*(4790) = 10.484, *p* < .001) bands joy contained significantly more energy than neutral utterances.

Conclusions

Results from acoustic analysis on the complete array of exemplars supported the study hypothesis which expected the acoustic profile of each state to differ on the selected acoustic parameters. Generally speaking, reported acoustic profiles coincide with the broader collective of literature of emotional and motivational prosody, with a few notable exceptions. In contrast to previously the reported acoustic profiles for autonomy-supportive and control, although findings indicated that voice quality played an important role in the unique expression of these qualities of motivation, the distribution was different to that previously reported (e.g., Weinstein, Zougkou & Paulmann, 2014; 2018). In this case, controlling utterances were not expressed with more energy in all frequency bands, but instead contained a greater proportion of high frequency energy and less low frequency energy than autonomy-supportive sentences. Also worthy of note, while it was expected that speakers intending to convey a controlling state would do so with a reduced pitch, the findings were the inverse; a slight pitch increase was found. Results suggest that speakers modulate their voices differently when intending to convey motivations and emotions, but there is no evidence of just one precise parameter that indexes the expression of either type of prosody. Thus, despite clearly being highly influential in the construction of distinctive acoustic profiles, there is no evidence to suggest that the expression of emotions is more reliant on voice quality than social intention, or in this case motivations are (c.f., Mitchel & Ross, 2013; Wickens & Perry, 2015). Instead it appears as though the expression of different prosodic messages is reliant on a complex combination of vocal cues. Nonetheless, importantly these findings not only reinforce the notion that different emotional states are communicated discretely, but they also suggest that motivations and emotions are expressed differently

through prosody, thus lend support to the idea that they are likely distinct communicative constructs. See Table 7 for a summary of parameter direction effects for each state.

Table 7: *Directional effects of parameters for each state in relation to neutral prosody.*

Parameter	Anger	Control	Joy	Autonomy
Pitch (normalised):				
Mean (F0)	>	>	>>	>
Range	>	=>	>	>
Amplitude (dB):				
Range	>	>	=>	=>
Speech rate	<	=	>	=
Energy Bands (Hz):				
0-500	<	<	<	<=
0-1000	<	<	<	<
500-1000	<	>	<=	<
1000-5000	>	>	>	<
5000-8000	>	>	>>	>

Note: < = Decrease; **>** = Increase; **<=/=>** = minor change; **<</>>** = large change; **=** = no change

Study 2

The previous study indicated that there are generalisable differences in how motivational and emotional tones of voice are vocally constructed. However, on their own these findings afford no insight into which cues lead to the effective inference of intended messages. Phrased differently, these encoding differences do not tell us anything about the patterns of cue modulation speakers use to effectively communicate their intended message, be it emotional or motivational. One reason for this is that by using a sample comprised of both, highly and non-prototypical exemplars, the identification of generalisable differences across states is possible (i.e., whether most speakers modulate a specific cue when intending to communicate a specific type of message). However, by not standardizing the quality of exemplars direct group comparisons (in this case comparison between states for a given encoder) are confounded and results could be the product of variation in stimuli across states (Matsumoto, 2002; 2007). Simply put, the use of an unscreened sample, did not consider that speakers may sometimes fail to convey their intended meaning and other times convey it perfectly and that this variation may imbalance the stimuli pool. The subsequent study set out to address these limitations by assessing the acoustic configurations of only the exemplars that were well recognised. It was predicted that for the recognised files, the states would differ on voice quality. In acoustically analysing files which effectively transmitted the intended message, this study reassessed the acoustic profiles presented in the previous study, this time accounting for the effective transmission of the intended message and provided an insight into which vocal cues listeners use to correctly infer motivation and emotional messages through tone of voice.

Method

Exemplar Selection

Judges

A total of 378 participants (287 Female; Mean age = 20.03, Range = 18-55, SD = 3.52; Mean years of education = 13.74, SD = 2.88) were recruited from the University of Essex Psychology Department to take part in the recognition study for course credits.

Exemplar presentation

Due to the extensive quantity of stimuli, audio files were split into 12 lists (6 containing 400 exemplars and 6 containing 399), from which Inquisit (version 5, 2016) randomly selected a subset of 300 for each decoder. Each subset was presented randomly, one at a time. A minimum of 30 decoders were recruited for each individual list, yielding a minimum a judgement count of 19 for any specific exemplar.

Procedure

In a forced choice paradigm, decoders were asked to categorise presented exemplars. In response to the question “How did the speaker sound?” judges were required to indicate from the provided options (“angry”, “pressuring”, “joyful”, “supportive” and “neutral”) how they felt the speaker sounded (i.e., what state the speaker was trying to communicate through their tone of voice). Prior to hearing any exemplars, participants were presented with descriptions of the categories (e.g., “Joyful: The speaker expresses happiness, joy or positive excitement”) and given explicit instructions to focus on the speaker tone of voice. Recognition sessions lasted approximately 35 minutes, performed online.

Acoustic Analysis

Exemplars recognised with 40% or better accuracy were selected for acoustic analysis. At this threshold, a total of 1740 exemplars were retained for analysis. See Table 6 for a breakdown per condition. In line with our previous study, the same analytic strategy was adopted here. As before, the same acoustic parameters were of interest, file amplitudes were normalised to a constant mean and pitch measures were normalised to reflect proportional changes.

Statistical analysis

As previously, separate mixed models with by-encoder and by-utterance random intercepts corresponding to state were used to analyse extracted acoustics, with Keppel's (1991) modified Bonferroni correction (p =. 02 for 10 contrasts) as the alpha for pairwise contrasts.

Results

Proportional distribution of recognised exemplars and parameter averages across states is summarised in Table 8.

Table 8: *Mean (and SD) of acoustic parameters for recognised exemplars across states.*

Parameter	Anger	Control	Joy	Autonomy	Neutral
Recognised exemplars:					
	194	200	237	292	817
Pitch (normalised):					
Mean (F0)	0.42 (0.26)	0.43 (0.18)	0.93 (0.39)	0.60 (0.29)	0.29 (0.19)
Range	1.63 (1.29)	1.14 (0.74)	1.54 (0.68)	1.31 (0.72)	1.09 (1.06)
Amplitude (dB):					
Range	42.06 (9.27)	37.84 (8.43)	34.33 (7.73)	35.47 (8.05)	34.01 (7.11)
Speech rate	0.24 (0.07)	0.22 (0.06)	0.18 (0.03)	0.19 (0.04)	0.19 (0.04)
Energy Bands (Hz):					
0-500	39.89 (2.43)	39.37 (2.57)	39.25 (2.72)	39.93 (2.31)	40.52 (2.39)
0-1000	38.65 (1.81)	38.65 (1.73)	38.60 (1.61)	38.79 (1.78)	39.15 (1.63)
500-1000	34.81 (4.71)	36.27 (3.75)	36.08 (3.63)	35.37 (4.33)	35.51 (3.58)
1000-5000	25.04 (3.95)	24.79 (3.85)	25.05 (3.99)	24.00 (3.97)	23.72 (3.80)
5000-8000	11.44 (5.13)	12.04 (5.11)	9.67 (6.41)	11.90 (5.55)	10.53 (5.51)

Pitch

Mean pitch (F0)

With regard to average pitch, analysis revealed a significant main effect of *state, F*(4, 1713) = 407.855, *p*<.001. Pairwise contrasts showed that all conditions significantly differed from each other (*p* <.003, *t* > 3.005, for all contrasts), with joyful utterances demonstrating the largest increase in pitch (m = .93, SD = .39).

Pitch range

A significant main effect of *state* was found for pitch range, *F*(4, 1713) = 18.128, *p*<.001. Elaboration of this effect revealed significant differences across the

majority of comparisons (all p <.011, t > 2.532), with the exception of differences between anger and autonomy, anger and control, anger and neutral and control and neutral (p >.033, t < 2.135 in these cases). Anger was expressed with the largest pitch range (m = 1.63, SD = 1.29). Findings are illustrated in Fig. 4.

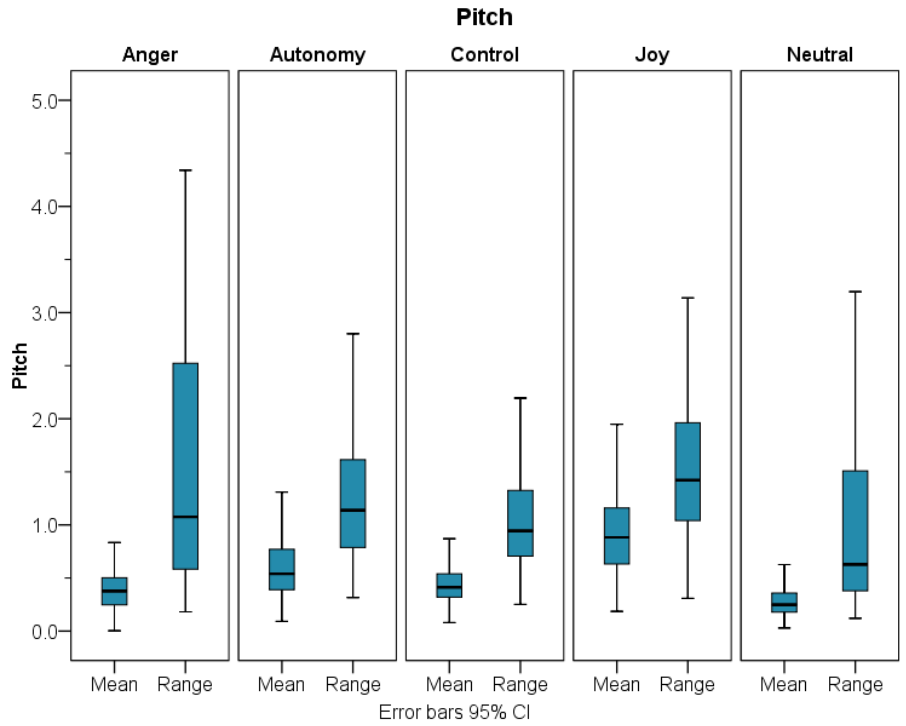

Figure 4: *Standardised mean pitch and range of recognised exemplars across states.*
Note: *Measures were normalised in reference to induvial speaker resting frequencies.*

Amplitude

Amplitude range

An effect of *state* was found with regard to amplitude range, $F(4,1681)$ = 27.957, p <.001. Pairwise contrasts demonstrated significant differences across all states (all p >.005, t < 2.791) with the exception of comparisons between autonomy and control, autonomy and joy, joy and control and joy and neutral (p >.021, t < 2.316 in all instances).

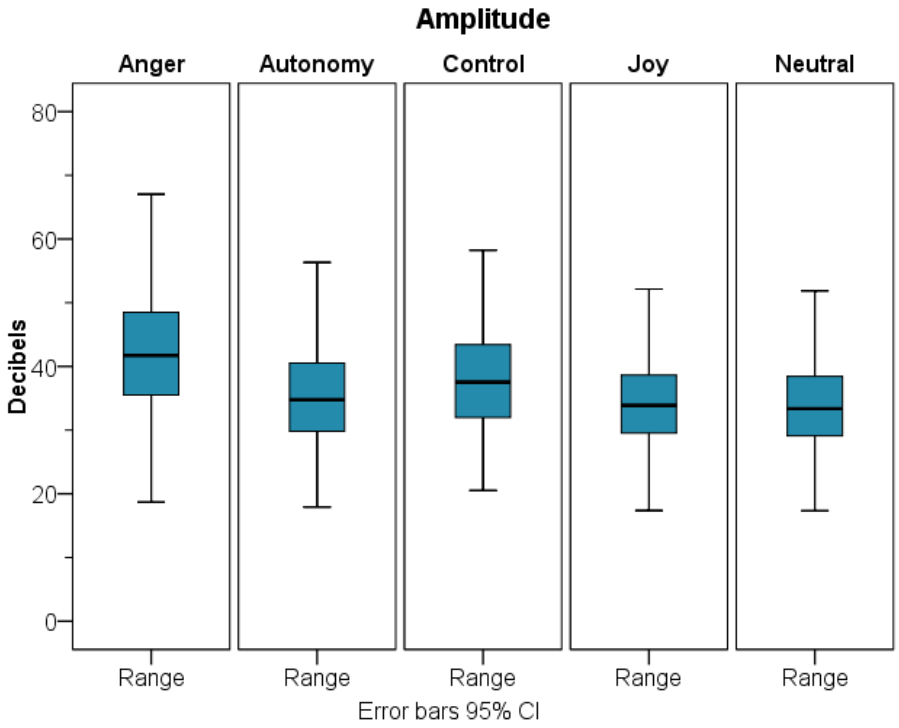

Figure 5: *Amplitude range for each investigated state for recognised exemplars.*

Speech rate

Duration per syllable revealed a significant main effect of *state*, $F(4, 1665) = 131.702$, $p < .001$. Effect elaboration revealed that only autonomy and neutral shared a similar speech rate ($b = .001$, $t(1666) = .660$, $p < .509$, whereas $p < .001$, $t > 5.536$ for all other contrasts).

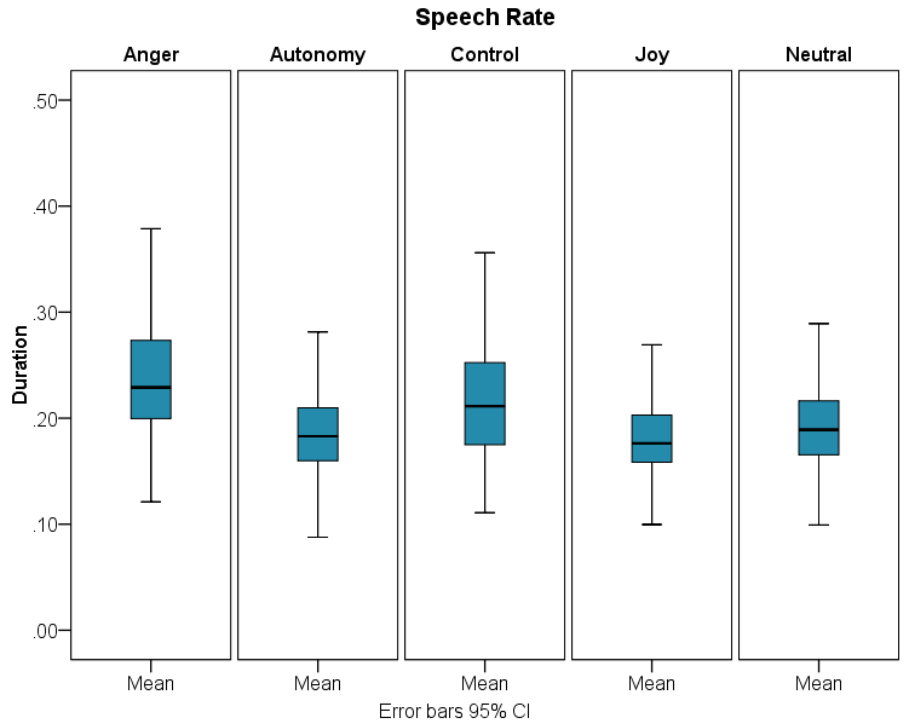

Figure 6: *Mean speech rate for each of the five investigated states.*
Note: *Higher durations indicate a slower speech rate (duration per syllable).*

Voice quality

Proportions of energy in measured bands of long-term average spectrum was found to differ as a function of *state* (0-500Hz, $F(4, 1693) = 19.329$, $p < .001$; 0-1000Hz, $F(4, 1702) = 11.914$, $p < .001$; 500-1000Hz, $F(4, 1675) = 2.946$, $p < .019$; 1000-5000Hz, $F(4, 1676) = 19.583$, $p < .001$; 5000-8000Hz, $F(4, 1689) = 12.911$, $p < .001$). As before, pairwise contrast will be presented individually across bands.

Anger – Autonomy

Angry utterances were expressed with significantly less energy than autonomy supportive communications in the 0-1000Hz band ($b = -.460$, $t(1708) = -2.906$, $p < .004$). No other energy distributions were significant ($p > .022$, $t < -2.287$ in all instances).

Anger – Control

Proportions of energy contained in the expression of these states only differed between 500 and 1000 Hertz (b = -.823, $t(1671)$ = -2.670, p < .008), with anger containing less energy in this band. Differences across other bands were non-significant (all p >.093, t < -1.682).

Anger – Joy

Angry exemplars contained less energy in all bands exceeding 500-1000Hz (In all cases p < .018, t >-2.364). Energy in 0-500hz and 0-1000Hz band was not significantly different between these states (all p >.113, t < 1.586).

Anger – Neutral

Angry sentences were conveyed with significantly less energy than every day neutral speech in bands up to 1000Hz (p < .001, t > -3.266 in all instances) and significantly more energy in the higher bands (1000-5000Hz and 5000-8000Hz, p < .010, t > 2.594 for both contrasts).

Autonomy – Control

Autonomy and control contrasts yielded no significant differences in energy across bands (p >.030, t < 2.172). A trending pattern was found in the 0-500Hz band, with autonomy containing less energy in this band (b = -0.409, $t(1710)$ = -2.172, p < .030).

Autonomy – Joy

Comparative to joyful, autonomy-supportive exemplars were expressed with more energy in the 0-500Hz, 0-1000Hz, 5000-8000Hz bands (all $p < .007$, $t > 2.683$) and less energy between 1000 and 5000 Hertz ($b = -1.197$, $t(1687) = -4.939$, $p < .001$).

Autonomy – Neutral

Autonomy-supportive sentences were expressed with less 0-500Hz ($b = -.346$, $t(1703) = -2.500$, $p < .013$) and more 1000-5000Hz ($b = .521$, $t(1682) = 2.774$, $p < .006$) energy than those in a neutral tone of voice.

Control – Joy

Controlling tones of voice were conveyed with a lesser concentration of energy between 1000 and 5000 Hertz ($b = -.828$, $t(1679) = -3.118$, $p < .002$), but more energy over 5000 Hertz ($b = 1.104$, $t(1694) = 2.639$, $p < .008$) than joyful expressions.

Control – Neutral

Compared to every day tones of voice, controlling messages were conveyed with a reduction in energy up to 1000 Hertz (0-500Hz and 0-1000Hz, $p < .001$, $t > -3.527$ in both cases), no difference between 500 and 1000 Hertz ($b = .006$, $t(1673) = .027$, $p = .979$), but more energy in the higher bands (all , $p < .001$, $t > 3.338$).

Joy – Neutral

Joyful utterances contained less energy in the 0-500Hz (b = -.1.132, t(1683) = -7.521, p < .001) and 0-1000Hz (b = -.603, t(1696) = -5.021, p < .001) bands. An increase in energy was found for joyful tones of voice between 1000 and 5000 Hertz (b = 1.717, t(1669) = 8.439, p < .001). Other contrasts were not significant (p >.709, t < .374).

Conclusions

Findings of the acoustic analysis of well-recognised exemplars indicated motivational and emotional messages are effectively communicated through different modulations of vocal cues, corroborating both, the findings of the previous study, and that the associated acoustic profiles are different from those of emotions. Analysis of the acoustic profiles linked with the effective communication of intended states (i.e., those recognised by listeners) highlighted that voice quality differences between the states was less pronounced compared to the previous study which used unscreened exemplars. This might suggest that although speakers generally modulate voice quality when intending to convey motivational and emotional messages through tone of voice, this modulation alone may not equate to the message being recognised and correctly inferred. In fact, similar to previous reports (e.g., Banse and Scherer, 1996) only average pitch accounted for enough of the variance to statistically distinguish between all states, independently of other vocal cues. However, to assume that listeners base their judgements and inferences on modulations of a single vocal cue is dismissive of the emotional prosody literature, which according to Paulmann (2015) implies that listeners rely on multiple acoustic parameters when detecting salience from emotionally rich auditory signals. A deeper inspection of these data supports this notion by indicating that the accurate inference of prosodic messages is likely

reliant on acoustic profiles comprised of varied cue configurations. For instance, while anger and control demonstrated similar levels of variability in their pitch, anger was expressed with a larger increase in average pitch, more loudness variability, a slower speech rate and a lower proportion of energy in the lower frequency bands than control. Across the board, every state differed from all others on a number or parameters, thus suggesting that when differentiation of a message is not possible based on a specific parameter, judges may turn to another vocal cue to assist them to decide. Thus, also suggesting that collectively these acoustic parameters can be uniquely configured to form distinct acoustic profiles that enable the accurate expression and recognition of motivational and emotional messages.

An alternative explanation for the variability contained across the acoustic profiles for target states is that listeners take the entire pattern of cues and add weights to each of the vocal cues contained within. This view may also explain why there is no distinct, single cue that enables the dissemination of motivational and emotional tones of voice, yet they can be recognised as different.

With respect to the generalised acoustic profiles found in unscreened or perhaps less prototypical exemplars, while there was some variation between the acoustic profiles (e.g., proportions of energy in the high and low frequency bands for joyful utterances, or a change in speech rate for controlling motivation), there was also a large overlap in their acoustic configurations (e.g., in both cases, joy was expressed with the largest increase in pitch, anger was spoken more slowly that other states, and control demonstrated the largest constraint in pitch variability). This overlap implies that some of these generalisable encoding differences, although perhaps do not guarantee the accurate inference, likely facilitate transmission of the intended message, hence why encoders utilise them when attempting to convey a

specific message through prosody. Interestingly however, it seems that changes in some cues for certain states (e.g., amplitude range and speech rate for anger) were more pronounced in the well-recognised sample, suggesting that the effective communication of these states may be highly reliant on those particular vocal cues.

However, that is not to say that when speakers intend to convey a particular message that they only modulate voice quality and neglect to modulate these other parameters (e.g., pitch, speech rate and amplitude range). In fact, these data simply suggest that although speakers modulate voice quality when conveying a message, that modulation on its own may not be sufficient for recognition and perhaps in some cases their modulations of the parameters that seem to be used by judges (e.g., pitch and speech rate) were too subtle to facilitate recognition. See Table 9 for a full summary comparison of parameter direction effects for each state across studies.

Table 9: *Acoustic parameter effects for states compared to neutral across studies.*

Parameter		**Anger**		**Control**		**Joy**		**Autonomy**
Study	1	2	1	2	1	2	1	2
Pitch:								
Mean (F0)	>	>	>	>	>>	>>	>	>
Range	>	>	=>	=	>	>>	>	>
Amplitude (dB):								
Range	>	>>	>	>	=>	=>	=>	>
Speech rate	<	<<	=	<	>	>	=	=
Energy Bands (Hz):								
0-500	<	<	<	<	<	<	<=	<
0-1000	<	<	<	<	<	<	<	<
500-1000	<	<	>	>	<=	>	<	<=
1000-5000	>	>	>	>	>	<=	<	=>
5000-8000	>	>	>	>	>>	<	>	>

Note: < = Decrease; **>** = Increase; **<=/=>** = minor change; **<</>>** = large change; **=** = no change; Study 1 used an unscreened sample and study 2 used only utterances there were recognised.

Study 3

Whilst collectively, the findings of the previous two studies indicates that motivations and emotions are conveyed differently and are recognised using different configurations of acoustic cues, how these constructs are perceived on-line still remains unanswered by these first two studies. Phrased differently, the previous studies suggest that motivational and emotional prosody are acoustically different which enables them to be differentiated through vocal cues but are limited in their ability to assess the distinctiveness of these forms of prosody from the perspective of listeners. To address this, the final study in this series explored how these states are processed in real-time. Guided by previous approaches to explore how emotions, attitudes (e.g., Paulmann & Kotz, 2008; Rigoulot, Fish & Pell, 2014; Wickens & Perry, 2015) and motivations (e.g., Zougkou, Weinstein & Paulmann, 2017) are processed, this study adopted an ERP methodology, with the expectation that all states would differ from each other at different processing time-points.

Considering that the primary purpose of this study was to assess the similarities and differences in the processing time-course of emotional and motivational prosody, comparisons of states were planned for each component of interest. Specifically, states were compared based on their likelihood of being misconstrued and compared to neutral. This yielded planned comparisons of two groupings; states with a negative (anger vs. control vs. neutral) and positive valance (joy vs. autonomy vs. neutral). These planned comparisons were comprised in this way to enable comparison of processing differences in response to subtle acoustic differences in the expression of expectedly distinct constructs (e.g., compared to neutral speech, control and cold-anger were both expressed increases in average pitch, increased variability in loudness, a reduction in speech rate than neutral, and

very similar energy distributions in the LTAS, but anger displayed an increase in pitch variability and control did not) . Anger is one of the most commonly studied emotional prosodies in the time-course literature and has reliably been shown to elicit relatively robust ERP components (e.g., Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Paulmann, Jessen & Kotz, 2012). Control, albeit limited to a single study, has been shown to elicit responses in the early salience detection and more in-depth processing stages and as such planned comparisons between control and anger are of the greatest interest. in contrast, the results for autonomy are less clear-cut and it will be insightful to see how it compares against happiness which has been shown to be differentiated from neutral in early and late components (e.g., Paulmann & Kotz, 2008; Paulmann et al., 2011)

Building on reports from the emotional, attitudinal and motivational prosody literature, it was hypothesised that motivations and emotions would differ in their processing pattern as early as the N1 component. Specifically, with it argued that sensory processing of frequency and intensity is processed within 100ms of sentence onset (e.g., Paulmann & Kotz, 2008), anger and joy are expected to contain sufficiently salient information (e.g., Liu et al., 2012) to elicit varied N1 responses, but autonomy-supportive and controlling prosody were not expected to differentiate this early (e.g., Weinstein, Zougkou, & Paulmann, 2017). Consequently, emotions and motivations are expected to differ in their N1 amplitudes. Motivations and emotions, particularly control and anger are expected to differentiate from neutral in the P200 component, but not from each other until a later point in time. With autonomy-supportive prosody reported to be spoken with less pronounced vocal cues (e.g., Weinstein, Zougkou, & Paulmann, 2017) and potentially lacking the

salient information which the N100 is sensitive to (Liu et al., 2012), it is predicted that joy and autonomy will differ in this early salience detection stage

Method

Stimuli Validation

Exemplar selection

Utterances from two encoders (one female aged 22 and one male aged 20) were taken from the initial encoding study based on experimenter decision prior to having undergone any acoustic analysis. Sentences were selected based upon recording errors (i.e., if a sentence from one encoder had to be rejected, it was not included for the other encoder). This selection process rendered a total of 600 utterances (60 utterances X 5 conditions X 2 Encoders) for validation. To avoid disparities in exemplar quality as a consequence of discrimination strategies used by judges (i.e., decoders may use a different strategy to distinguish exemplars in this study taken from 2 speakers, compared with from 14 speakers in the previous studies) an independent validation study was deemed necessary.

Judges

Fifty native English speakers (36 Female; Mean age = 27.06, Range = 18-55, SD = 11.79; Mean years of education = 14.24, SD = 3.16) were recruited from the University of Essex and social media (e.g., Facebook) to take part in this validation study online. For their time, participants were entered in a prize draw for a £20 Amazon Voucher.

Exemplar presentation

The 600 exemplars were pseudo-randomly split (i.e., the only requirements were that each list contained 30 audio files from each encoder per condition) into two

lists of 300. Using Inquisit (2016), lists were presented, one exemplar at a time. 25 decoders were allocated to each list.

Procedure

Following the same validation process as in the earlier study, decoders were asked to categorise presented exemplars. In response to the question “How did the speaker sound?”. In a forced choice paradigm, decoders could choose between “angry”, “pressuring”, “joyful”, “supportive” and “neutral”. As before, category descriptions and explicit instructions to focus on the speaker tone of voice were given prior to stimuli presentation. Recognition sessions were performed online and like in Study 2, lasted approximately 35 minutes.

ERP Study

Participants

Thirty-eight native English speakers (21 Male; Mean age = 21.03, Range = 18-32, SD = 2.75) were recruited from the University of Essex. All participants were right hand dominant, assessed by an adapted version of the Edinburgh Handiness Inventory (Oldfield, 1971). Participants were rewarded with their choice of £10 or 2 course credits for their time.

Stimuli

Following stimuli validation, 16 well recognised exemplars (i.e., recognised with over 40% accuracy) were selected for each condition from each speaker. For conditions in which 16 utterances were not recognised with above 40% accuracy (control for both encoders and anger for the female encoder), 8 recordings that did meet the recognition accuracy criteria were repeated. This process was mirrored for both encoders (i.e., in the event one of the encoders provided 16 recordings at the required quality but the other encoder did not, 8 exemplars were repeated for both

encoders for that condition). In total, 128 recordings were selected for inclusion (16 for anger, 16 for control, 32 for autonomy, 32 for joy, 32 for neutral).

Because pitch, loudness and speech rate have previously been linked to the moderation of salience detection (e.g., Pantev, Elbert, Ross, Eulitz & Terhardt, 1996; Picton, Woods, Baribeau-Braun & Healey,1977; Chang, et al., 2018), and so far in this investigation have presented themselves as key parameters for understanding how motivational and emotional tones of voice may differ, stimuli selected for use in this study was acoustically analysed on these parameters (using Praat). Consistent with the previous studies in this investigation, amplitudes of stimuli were normalised to a constant mean and pitch was acoustically measured in relation to proportional changes (i.e., changes from encoder resting frequencies).

Acoustical analysis of stimuli yielded similar acoustic profiles to those established in the previous studies, with the notable exceptions of reductions in pitch variability for control and autonomy, which in the previous studies were found to increase compared with neutral. This change may be a result of a large reduction in idiosyncratic differences within the sample (i.e., previous stimuli pools contained utterances from 14 encoders, whereas only recordings from 2 encoders were selected as stimuli for this experiment, thus more speakers would likely result in more variation in pitch). Overall, following the same pattern established previously, stimuli differed in pitch parameters. Average pitch differed across all conditions. Compared with neutral everyday speech, joyful exemplars were conveyed with the largest increase in pitch (m = 1.26, SD = .34) and controlling messages included the lowest increase in mean pitch (m= .50, SD = .20). With respect to pitch variability, controlling utterances displayed the most constrained pitch (1.63, SD = 1.25), whereas the most variability in pitch was associated with angry messages (2.09, SD

= 1.44). Likewise, stimuli differed with regard to variability in intensity (joy varied the least, m = 33.51, SD = 5.11; control contained the most variability in amplitude, m = 37.11, SD = 6.53) and speech rate. As has been the case through-out, anger was expressed with the slowest rate of articulation (see Table 10 for acoustic summary of included stimuli).

Table 10: *Results from acoustical analysis of included stimuli for all tested conditions.*

Parameter	Anger	Control	Joy	Autonomy	Neutral
Pitch (normalised):					
Mean (F0)	0.58 (0.24)	0.50 (0.20)	1.26 (0.34)	0.77 (0.25)	0.40 (0.29)
Range	2.09 (1.44)	1.63 (1.25)	1.79 (0.82)	1.71 (0.97)	1.76 (1.48)
Amplitude (dB):					
Range	37.86 (8.67)	37.11 (6.53)	33.51 (5.11)	34.43 (7.63)	34.10 (5.35)
Speech rate	0.20 (0.03)	0.19 (0.04)	0.19 (0.04)	0.19 (0.03)	0.19 (0.04)

Procedure

EEG recordings were acquired in a sound attenuated booth, in which participants were seated approximately 100 cm from a computer screen. Materials were presented in four blocks of 40 trials in a completely random order using SuperLab 5 (2015). Trials comprised of a fixation cross in the middle of the computer screen for 300ms, followed by a vocal stimulus (average duration was 1200ms, with a range of 770-2200ms) via speakers on both sides of the monitor and ended with an inter stimulus interval of 1500 milliseconds. Although no task was given, to promote participant engagement with the materials, they were asked to listen carefully and told that they would be asked simple questions in relation to what they had heard at the end of the session. Prior to the experimental blocks, five practice trials and two “yes vs. no” questions were presented to familiarize participants with the procedure and also to simulate the expected study design (i.e., that there would be simple comprehension questions at the end of the session).

EEG recording

This experiment mimicked the recording strategy utilised by Zougkou, Weinstein & Paulmann (2017). EEG was measured using a custom-made cap (waveguard) with 63 mounted Ag-AgCl electrodes, according to the modified 10-20 system and an ANT amplifier (72 channel Refa). Electrode resistances were kept below 20KΩ in all cases, CZ served as ground electrode and the reference electrode was placed on the left mastoid. Data was re-referenced offline to the averaged mastoids. Signals were continuously recorded using a band pass filter between DC and 102Hz and were digitised at a 512 Hz sampling rate. Eye movements were recorded for artefact rejection purposes using disposable Ambu Blue Sensor N EEG Electrodes positioned above, below (vertical EOGs) and on the left and right (Horizontal EOGs) of the participants eyes

Data analysis

Data were filtered offline using a band pass filter (0.01 – 30 Hz) and a baseline correction was applied. The mean of the baseline time window (-200-0ms) for each ERP channel was subtracted from the averaged signal of that channel. All trials containing muscle or EOG artefacts above 30.00 μV were automatically rejected using EEProbe software, after which data was visually inspected to exclude trials that contained additional artefacts and drifts. Upon data inspection, of the thirty-eight recorded participants, seven were excluded due to insufficient data points, yielding less than 15 useable data points for any state (see Appendix 4 for full data point summary). The remaining sample comprised of thirty-one participants (18 Male; Mean age = 21.06, Range = 18-32, SD = 2.92), from which 23% of trials were rejected (range for different conditions 22% - 24%). Subsequent to data cleaning, ERPs from individual electrode-sites were averaged for each condition for each

participant. Averages contained a 200ms pre-stimulus baseline and epochs lasting 800ms post stimulus onset, which were time locked to sentence onset of stimuli.

Because the primary aim of this study was to ascertain whether motivations and emotions are processed with different time-courses, despite motivational prosody being reported to not elicit differential N1 ERP amplitudes (Zougkou, Weinstein & Paulmann, 2017), because this component has been linked with emotional prosody (e.g., see Paulmann, 2015 for review) it was still of interest to this study. Time window selection was informed by visual inspection of the data and previous research and thus the window for the N1 component was set between 90-150ms, the P200 was set between 170 and 220ms, the later negativity was explored between 350-600ms. Since late long-lasting effects have been linked to some attitudes (e.g., sincerity; Rigoulot, Fish & Pell, 2014) the present study also explored late long-lasting potentials between 500 and 800ms. With motivational prosody processing patterns limited to the study by Weinstein, Zougkou and Paulmann (2017), electrode sites were grouped in the same fashion; by hemisphere (left and right) and region (frontal, central and parietal). Grouping yielded the following seven regions of interest (ROIs): right frontal (F6, F4, FC6, FC4); left frontal (F5, F3, FC5, FC3); right central (C6, C4, CP6, CP4); left central (C5, C3, CP5, CP3); right posterior (P6, P4, PO8, PO4); left posterior (P5, P3, PO7, PO3); and midline (Fz, Cz, CPz, Pz).

Separate repeated-measures analyses of variance (ANOVAs) were used to analyse the mean amplitudes for each time window and respective ROIs. In all analyses, ROI and *state* were treated as within-subject factors, with the Greenhouse-Geisser correction applied where required. Significant main effects and interactions involving *state* at $p < .05$ were followed up with pairwise comparisons. Planned

comparisons between control, neutral and anger as well as autonomy, joy and neutral were also conducted regions of interest identified by the previous literature and midline electrodes to inform the primary aim of this study. For consistency through-out, post-hoc contrasts and planned comparisons were conducted using Keppel's (1991) modified Bonferroni correction, which resulted in pairwise comparisons at p <.02 (for 10 contrasts and 4 test condition *df*) and p <.03 for planned comparisons (in which there were 3 contrasts and 2 *df* associated to the test condition). Contrasts significant at non-corrected alpha (p<.05) will be reported to inform readers of emerging patterns.

Results

N1 (90-150ms)

Primary analysis of this very early window revealed no main effect of *state* (F(4, 120) = 15.927, p = .727) or interaction with ROI (F(8.56, 256.93) = .460, p = .625). Planned comparisons revealed no significant *state* differences in frontal ROIs (all p > .216) or in the midline ROI (p >.261, in all cases). Thus, findings in this time window provide no evidence for any differences in ERP amplitudes in response to motivational or emotional prosody in this very early component.

P200 (170-220ms)

Analysis at this component yielded no main effect of *state* (F(4, 120) = 8.337, p = .887), but a significant interaction between *state* x ROI (F(8.43, 252.80) = 6.714, p = .008). Pairwise comparisons warranted by this interaction revealed an enhanced positive ERP amplitude for anger compared to neutral that approached significance (p = .045, 95% CIs [.027, 2.121]) in the left frontal ROI. Looking at midline electrode sites, planned comparisons revealed significantly smaller P200 amplitudes for

control when compared to neutral p = .018, CIs [-2.308, -.230]). Amplitude differences between joy and neutral approached significance at midline electrodes, with joy eliciting a less positive component than neutral (, p = .052, CIs [-2.423, .010]). Other planned comparisons were non-significant (p >.075 in all instances). Results of this time-window confirm that neutral and angry prosody can be distinguished fairly early after sentence onset, an effect predominantly found at left frontal electrode sites. In addition, a similar early differentiation is found between control and neutral; however, this effect is more centrally located. There was no indication that controlling and angry prosody were differentiated from each other at this point in time.

Late negativity (350-600ms)

With respect to the late negative component, primary analysis yielded no main effect of *state* (F(2.09, 62.846) = .598, p = .560) or an interaction (F(9.117, 273.50) = .429, p = .921). Planned comparisons revealed no effects that approached significance in this window of interest (p >.132 for all contrasts). In short, motivational and emotional prosody seems to follow a similar processing time-course at this stage in time.

Late long-lasting potential (500-800ms)

No main effect of *state* (F(2.14, 64.20) = 1.079, p = .349) or interaction with ROI (F(8.07, 242.14) = .649, p = .737) was found in the primary analysis. Planned post-hoc comparisons in the midline ROI revealed an enhanced positive going amplitude for joy compared with autonomy- supportive that approached significance (1.336, p = .041, CIs [.058, .2.615]). No other contrasts approached significance in

this window (p >.094 in all cases). In sum, there is an indication that autonomy-supportive and joyful prosody can be differentiated from each other within this time-window of interest.

Conclusions

The final study of this investigation explored the differences and similarities in time-course and neural resources associated with emotional and motivational processing. The results demonstrated some evidence for the differentiation of emotions and motivations compared with neutral prosody as quickly as 200ms onset as we expected. No evidence was found for the anticipated differentiation between emotions and motivations or from neutral speech at the N1 component. Similarly, no differences were found for the late negative component, suggesting that both motivational and emotional messages underwent similar evaluation processes at this point in time. Perhaps more interestingly, there was some weak evidence of potential differentiation between emotions and motivations, in particular joy compared with autonomy-support between 500-800ms. Conversely to Mitchell & Ross's (2013) hypothesis, our results only provide support for the possibility that emotions and motivations share the same underlying time-course but may vary slightly in neural networks responsible. Because only weak evidence of differentiation between emotions and motivations was found in this study, similar to Wickens & Perry (2015), this study suggests that emotions and motivations are processed at a similar point in time in the brain (c.f. Kotz & Paulmann, 2011).

General Discussion

Building on work studying motivational (e.g., Weinstein, Zougkou & Paulmann, 2014; In Press; Zougkou, Weinstein & Paulmann, 2017; Paulmann, et al., 2018) and emotional prosody (e.g., Banse and Scherer, 1996; Sobin & Alpert, 1999; Paulmann & Kotz, 2008; Pell, et al., 2009; Paulmann, Ott & Kotz, 2011) the present series of studies were the first to directly compare these forms of prosody. In identifying similarities and differences between emotional and motivational prosody on an encoding and decoding level, this investigation has comprehensively demonstrated that these forms of prosody are in fact uniquely different and as such are deserving of their own respective literatures.

By firstly establishing the acoustic profiles associated with the target emotional and motivational states, how each of these states sound in comparison to each other was shown (i.e., how their associated tones of voice are acoustically different from one another). Moreover, these differences were not only assessed and presented with regard to generalisable production differences (i.e., the general tone of voice fluctuations speakers use to convey the intended state), but also with respect to the vocal cues that lead to these states being accurately transmitted and inferred through tone of voice. The findings illustrate that motivational and emotional messages are produced with different configurations of vocal cues and as such are communicated through distinctly different tones of voice. The next study suggested that although it is possible that motivational and emotional tones of voice are differentiated in the brain during re-analysis, they in fact share a similar processing time course and partly overlapping neural resources.

Acoustic characteristics of motivational and emotional prosody

Acoustically, these findings lend support to the notion that motivational and emotional prosody are distinct from each other and as such should not be conflated. Despite some evidence of greater speaker reliance on voice quality modulations with respect to motivational messages, when taking into account whether the intended message was effectively transmitted, our data suggests that the expression and inference of neither construct is more or less reliant on voice quality when latter is measured as energy distributed in different frequency bands (c.f., Mitchell & Ross, 2013; Wickens & Perry, 2015). Other researchers have used other voice quality indicators, such as Harmonics-to-noise ratio (HNR; Yumoto, Gould & Baer, 1982; e.g., Cheang & Pell, 2008) and shimmer or jitter (e.g., Wolfe, Fitch & Cornell, 1994; Rabinov, Kreiman, Gerratt, & Bielamowicz, 1995), thus this hypothesis could be explored in more detail by future research using a different indicator of voice quality.

Whilst the present findings reinforce the well documented importance of pitch modulations in the prosodic expression of expression of emotions (e.g., Scherer, et al., 1991; Banse & Scherer, 1996; Paulmann & Uskul, 2014) and motivations (e.g., Weinstein, Zougkou & Paulmann, 2014; In Press), they also clearly demonstrate that the effective communication of emotional and motivational states through tone of voice relies on the modulation of more than one vocal cue.

In line with previous assertions that listeners rely on more than a single cue to detect emotionally important information contained in a vocal message (Paulmann, 2015), results of acoustic analysis on well recognised stimuli suggested that the effective communication of intended states via prosody is achieved through acoustic

profiles comprised of a unique configuration of vocal cues. For instance, although cold-anger and control only demonstrated minor differences in voice quality (indexed by proportions of energy in different frequency bands), they differed on average pitch, amplitude range and speech rate. Similarly, although control did not differ from neutral in pitch variability, it was expressed with an increase in average pitch, more loudness variability, a slower speech rate, a reduction in low frequency energy and more high frequency energy. Simply put, every investigated state was differentiable from all others on at least 3 of the extracted acoustic parameters, thus indicating that cold-anger, joy, autonomy-supportive and controlling tones of voice are comprised of a unique configuration of vocal cues, supporting the notion that emotions and motivations are conveyed differently through prosody. More holistically, these findings enable the assertion that even though they share some vocal cues in common, motivational and emotional tones of voice are constructed with different cue configurations. This is hardly surprising given that via one of these constructs' speakers have the intent to motivate someone to action but through the other they have the intent to indicate to the listener how they feel. Therefore, as these constructs differ in what they are used to achieve and are acoustically different, it seems reasonable to argue that they are distinct from each other and conflation of these constructs is unwarranted.

Interestingly, there was some indication of more pronounced pitch, intensity and durational cues in the well-recognised over the unscreened sample for emotions but not motivations (see Table 8 for summary). This is interesting because these parameters have been reported to be influential in the detection of salient information in emotional messages (e.g., Picton, et al.,1977; Pantev, et al., 1996; Chang, et al., 2018). Although far more extensive investigation is required, this observation implies

that it is possible that motivational prosody is less reliant on high intensity (or highly pronounced) vocal cues than emotions; an idea that gains some support from research suggesting that expression of attitudes relies on more *intentionally controlled* (as opposed to involuntarily produced) processes which may cause them to diverge from emotional expressions (Mitchell & Ross, 2013) and that speakers rely on more subtle and varied vocal cue manipulations when conveying motivations than emotions (e.g., Weinstein, Zougkou & Paulmann, 2014).

These more pronounced vocal cues to a large extent support and are in line with the previously reported acoustic profiles for joy and cold-anger (e.g., Scherer, et al., 1991; Banse & Scherer, 1996; Sobin & Alpert, 1999). However, conversely to Banse & Scherer (1996), our findings demonstrated that speakers greatly reduced their speech rate when conveying cold-anger, but this might be a consequence of language differences or cultural display rules and expectations (e.g., see Matsumoto, et al., 2002; Pell, Monetta, Paulmann & Kotz, 2009, for more information about cultural constraints on emotional expressions).

Generally speaking however, the acoustic profiles obtained in this investigation are in line with the previous established cue configurations, in the sense that joy was characterised by more variation in pitch and was spoken faster. Likewise, anger, or in this case cold-anger was expressed with a mild increase in average pitch and more fluctuations in loudness. However, conversely to Sobin and Alpert (1999), here joy was expressed with a large increase in average pitch. The same was observed for motivational tones of voice; some acoustic cues lined up with those reported by Weinstein, Zougkou, and Paulmann (2014, 2018), but others were the inverse. To be more precise, while autonomy-supportive messages were expressed with a higher average pitch and in a less harsh tone of voice, unlike

previous findings our data indicates that autonomy-supportive utterances are expressed faster than controlling messages. Interestingly, however the reported reductions in loudness variation for autonomy-supportive utterances was only found in the unscreened sample. The inverse was actually found when accounting for recognisability of the conveyed messages. This suggests that although controlling tones of voice are frequently expressed with more variability in their loudness parameter, this cue does not equate to the message being recognised. Phrased differently, when intending to communicate motivations speakers may modulate this vocal cue, but our data suggests that this cue may not actually carry the message and thus may in fact be more related to intention than the effective motivation of others. Considering the cultural rules and expectations that constrain the expression of emotions, it seems likely that social communicative functions are bound by similar cultural and social expectations. As such, perhaps certain levels of loudness are deemed socially acceptable when trying to be supportive, but the sense of support is possibly conferred by different vocal cues, such as the harshness of the voice or pitch or even how cues are modulated in relation to each other (e.g., it might be unacceptable to communicate support with have a very high pitch and a very loud voice, but may be ok to use low pitch and loud voice or high pitch and less loud voice).

Furthermore, differences in parameter variability across states and between screened and unscreened exemplars indicates that there is a large amount variation across speakers in the way they communicate these states. In fact Weinstein, Zougkou and Paulmann (2018) reported between how students and actors modulated pitch when conveying autonomy-supportive sentences. Given that communications of these constructs differ in their intended purposes (i.e.,

motivations provoke others to act, whereas emotions indicate how the speaker feels), it is important for us to understand how these messages are conveyed and subsequently inferred and disseminated. According to Paulmann, et al. (2016) the misinterpretation of vocally expressed emotions can have a detrimental impact on social interactions and can lead to issues such as social exclusion of the listener or speaker. In a similar vein, not being able to disentangle motivational messages in which the listener is being called to action, from messages indicating the feelings of the speaker could negatively impact social reciprocity and associated functionality (e.g., effectively working together). Also, from an empirical point of view, with more recent studies investigating factors that potentially mediate the transmission and inference of vocally communicated emotions (e.g., Paulmann, et al., 2016; Uskul, Paulmann & Weick, 2016), the importance of knowing the acoustic profiles associated to these communications is extremely important. In order for research to effectively assess the more fine-grained aspects of prosodic encoding and decoding a solid acoustic foundation is required.

Following arguments made by Matsumoto (2002; 2007) in this investigation we assessed the differences and similarities between motivational and motivational prosody using exemplars of similar quality. By ensuring not only the use of exemplars of equal quality, but also an identical baseline for all states, the present investigation convincingly demonstrates that motivational and emotional tones of voice are, on an acoustic level distinct. This research was the first to directly compare the acoustic profiles of these conceptually different forms of prosody and in doing so has contributed to the wider literature as well as each of the respective literatures. Cold-anger in general has been studied less frequently than hot-anger and for English stimuli joy has been heavily neglected as most studies have selected

happiness as the positive emotion in this language (e.g., Pell, et al., 2009; Paulmann & Uskul, 2014; Paulmann, et al., 2016). Consequently, the acoustic profiles established in this investigation contribute to those neglected in the emotion literature. With respect to motivations, although Weinstein, Zougkou and Paulmann (2014; 2018) clearly demonstrated how controlling and autonomy-supportive tones of voice differ compared to each other, they offered no insight into the modulations respective to neutral every day speech; a deficit this research sought to rectify. On the grander scale, contributions are also made to a number of ongoing debates. Of primary concern is the conflation of these constructs in the literature. The results of our acoustical analysis indicate that this practice is unwarranted and motivational and emotional tones of voice should have their own respective literatures. In addition, by analysing and identifying differences in unscreened as well as screened exemplars, this research suggests that the use of unscreened samples is potentially not suitable when investigating the “successful” communication of intended prosodic messages (c.f., Bachorowski & Owren, 2008).

Processing time course of motivational and emotional prosody

Informed by time-course investigations into emotions, attitudes and motivations (e.g., Paulmann, Ott & Kotz, 2011; Rigoulot, Fish & Pell, 2014; Wickens & Perry, 2015, Zougkou, Weinstein & Paulmann, 2017), four different processing stages were focused on: Processing of sensory information and vocal cue extraction (N1), differentiation and evaluation of saliency cues (P200), more effortful analysis of meaning (later negativity), and (re)analysis of prosodic information (late long lasting potential). Examination of these different stages permitted the effective description and comparison of the time-course underlying vocal motivational and emotional signal processing. Whilst it was observed that control and anger differentiated from

neutral within 200ms within different regions, the lack of later differentiation between these states suggests that they also share processing steps and resources. Interestingly, this effect for “negative” or potentially “withdrawal” states differs from the findings for more “positive” or “approach” states. Specifically, the comparison between joy and autonomy-support when looking at a later more in-depth focus stage (500-800ms) turned out to be significant. Thus, taken together, these data nicely demonstrate both similarities and differences in processing of motivational and emotional prosody; they also nicely highlight that positive and potentially more negative social intentions as communicated through voice warrant separate analyses. Findings for each component will be discussed separately below.

Processing of sensory information (90-150ms)

This component has not been reported to be a point of differentiation for motivational tones of voice (i.e., between autonomy-supportive and controlling tones of voice; Zougkou, Weinstein & Paulmann, 2017); however, some limited studies have demonstrated differing N1 amplitudes in response to emotional prosody (e.g., Liu et al., 2012; Pell, Rothermich, Liu, Paulmann, Sethi & Rigoulot, 2015). If true that motivations are not differentiated from neutral communications this early while emotions are, this component should be of interest as an expected point of differentiation between emotional and motivational tones of voice. Contrary to this view, the current data provide no evidence of differentiation in this very early processing stage. There is some evidence that N1 responses are modulated by the saliency of provided information (Liu et al., 2012) as well as selective attention mechanisms (Hillyard, Hink, Schwent & Picton, 1973). As such a potential explanation for the lack of N1 differentiation on the measures examined, particularly between emotional and neutral stimuli, could be that despite explicit instructions to

focus on how speakers sounded and being told that there would be some questions about what was heard at the end of the study the lack of active task to direct attention to the stimuli may have led to less attention being allocated to the saliency cues required to elicit this component. Perhaps scientifically less interesting (but potentially more likely), is the possibility that previous studies reporting emotional and neutral differentiation relied on vocal samples that differed in amplitude. Traditionally, the N1 component has been associated with loudness processing (e.g. Picton & Hillyard, 1988). Thus, in studies where stimuli have not been normalised with regard to amplitude, it is more likely to find N1 differences between categories that differ in loudness. Here, we presented stimuli that were normalised with regard to average amplitude and thus may have failed to find differentiation between emotions and motivations because previously reported effects are predominantly driven by these stimuli differences.

Evaluation of saliency cues (170-220ms)

In line with the study hypothesis, motivations and emotions (specifically control and anger) differentiated from neutral, but not each other in the P200 component. Results showed enhanced P2 amplitudes at left anterior electrode sites when comparing angry and neutral prosody. In contrast, controlling and neutral prosody differentiated at separate electrode-sites, specifically more centrally located ones. Here, we found smaller P2 components for control compared to neutral prosody. There was also some limited indication that joy and neutral were differentiated at electrode-sites in the same central region. Overall, the findings are in line with our expectations and provide some evidence to support the notion that motivational and emotional prosody are mediated by partly different neural sources given that effects are found at different electrode-sites. Future studies using source

localisation methodologies will have to confirm this notion; for now, the important contribution that the current data make to the literature is that both motivational and emotional prosody can be differentiated from neutral prosody within 200ms after stimulus onset, adding to the growing literature that different vocal social signals are initially scanned for saliency by the listener (e.g., Regal, Gunter & Friederici, 2010; Kotz & Paulmann, 2011; Iredale, et al., 2013; Schirmer, et al., 2013; Rigoulot, Fish & Pell, 2014).

The distribution of the current effects is in line with that of Liu and colleagues (2012) who reported P200 reductions in the central compared with frontal and frontocentral regions. In addition, evidence suggests that this positivity may be sensitive to distributional changes relative to task demands and speaker effects (e.g., Kotz & Paulmann, 2008; Chen, Zhao, Jiang & Yang, 2011; Wickens & Perry, 2015). For instance, Kotz and Paulmann (2008) showed P200 distribution differences when reporting data for a probe-verification vs prosodic classification task. Similarly, Chen and colleagues (2011) reported a bilaterally distributed effect for a prosodic classification task and a more mid-left lateralised effect when testing prosody implicitly (i.e. during a probe verification task). Here, we applied no task and thus also asked our participants to process prosodic attributes implicitly; the data pattern that emerged is thus nicely mirroring that of Chen et al. (2011). In fact, it has been argued that neural resources responsible for saliency processing of prosody may not be fixed, but instead may depend on allocation of attention as a result of the task at hand (e.g., Wildegruber, Hertrich, Riecker, Erb, Anders, Grodd & Ackermann, 2004; Wickens & Perry 2015). This is in line with reports that show how prosody processing associated ERP patterns can be influenced by differing task demands

(e.g., Astésano, Besson & Alter, 2004; Schirmer, Kotz & Friederici, 2005; Kotz & Paulmann, 2007; also see Paulmann, 2015, for discussion).

Nonetheless, seeing as though control and anger both differentiated from neutral at this component, but not each other, we can infer that both forms of prosody may contain enough salient information to be distinguishable from neutral, everyday tones of voice within 200ms after stimulus onset. Furthermore, the differently distributed P200 effects for control and anger compared with neutral indicate the possibility that partly differing neural structures may be in operation during differentiation of these constructs at this stage. Similar to Weinstein, Zougkou and Paulmann (2017), in response to non-biasing sentences, control was shown to differentiate from neutral in the midline at this component, however unlike their findings, we found no differentiation between autonomy-supportive and controlling tones of voice in either region that they reported effects (e.g., frontal and midline). Although as the effect sizes between autonomy-supportive and controlling prosody were relatively small (.25-.26) and no effects were reported for autonomy-support compared with neutral, it is possible that the differences observed between these studies is a result of methodological and/or stimuli differences. Thus, taken as they are, our findings suggest that these types of prosody share a similar processing time course, but may differ slightly in neural networks.

Analysis of meaning (350-600ms)

Despite the motivational prosody literature (Zougkou, Weinstein & Paulmann, 2017) reporting more positive going amplitudes between control and autonomy-supportive prosody, but not neutral between 350-600ms onset, the present investigation found no evidence of differentiation at this point in time between any state or from neutral. This is perhaps less surprising when looking at effects of

semantic content in these contexts. In Zougkou et al. (2017), stronger effects were generally found for semantically biasing sentences as opposed to semantically neutral sentences. Here, we used similar non-biasing sentences (e.g., “why don’t you ask for help?”) making it a bit more difficult to find prosodic evaluation effects (with implicit task instructions). In fact, in the past, effects around this time-window, such as the well-studied N400, are extensively linked to the integration and evaluation of semantic information (e.g., Van Petten & Kutas, 1988; Van Petten, Coulson, Rubin, Plante & Parks, 1999; Wambaq & Jerger, 2004; Schirmer, Kotz & Friederici, 2002, 2005; Van Petten & Luka, 2006; Kotz & Paulmann, 2007). More importantly however, speech processing, has been shown to rely on the integration of verbal like semantics or syntax and non-verbal information such as prosody (e.g., Steinhauer, Alter & Friederici, 1999; Eckstein & Friederici, 2006; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008; Paulmann, Jessen & Kotz, 2012). Some evidence suggests that when listeners focus on linguistic prosody they are unable to ignore semantic information, but prosodic information can be ignored when semantics are the target of their attention (e.g., Besson, Magne, & Schön, 2002). Assertions similar in nature have also been made in the emotional prosody literature, with it being argued that semantic processing may override prosody processing when presented together (e.g., Kotz & Paulmann, 2007). Applied directly to our results, focus on processing semantic information may explain why all conditions elicited a similar negativity and thus may explain why no differentiation was observed in this time-window. However, with that being said, Zougkou, Weinstein and Paulmann (2017) reported effects between autonomy supportive and controlling tones of voice at this point in response to semantically valid sentences. They found small to medium positive going effects for control compared with autonomy-supportive (.37 in frontal

regions and .41 at midline), but not compared with neutral. In contrast, however the semantically valid sentences used in their study had little or no contextual saliency (e.g., “You are quite tall for your age”), whereas it is possible that contextual utility of the sentences used in the present investigation (e.g., “I suggest you don’t rush this”) may have conferred additional information via implicit associations with specific settings. Supporting this possibility, when Zougkou, Weinstein and Paulmann (2017) assessed processing patterns associated with linguistic and prosodic content the effects were larger (.52 and .45) and were found compared with neutral as well. Consequently, despite previous findings demonstrating differentiation between control and autonomy support at this processing time-window, our findings demonstrate no such effect, thus suggest that these constructs when carried via semantically valid and contextually relevant sentences share the same processing pattern at this component.

Later analysis of prosodic information (500-800ms)

Trending differences between joy and autonomy-support indicate the possibility that positive emotions and motivation may undergo slightly different evaluations in this later point in time as predicted by Mitchell & Ross (2013). Various later components have been linked to either the re-analysis of prosodic information or the more focused processing of prosodic information, such as the closure-positive shift (CPS; e.g., Steinhauer, Alter, &Friederici, 1999) and the P800 (e.g., Astésano, Besson & Alter, 2004), for instance. Therefore, if we take this later component to reflect further, more focused analysis of prosodic information contained in the message, it seems to indicate that information more directly relevant to the listener is preferentially processed. That is, when presented with information that indicates that the speaker is happy or information that confers a sense of feeling supporting the

action the listener is about to perform, supportive information, which is of direct relevance to the listener takes precedence.

Therefore, the ERP amplitudes associated with joy and autonomy-supportive at this point in time could well indicate that the stimuli associated to these conditions was analysed differently. Given that these states were conveyed with similar acoustic configurations, it is not surprising that listeners only start to differentiate between these states at this later point in time when cognitive resources can focus on thoroughly analysing what the speaker's social intention is. However, if true that positive emotions and motivations are disentangled prosodically in this later processing stage, the question remains why similar results were not found for negative emotions and motivations (here control and anger)? Although surprisingly limited, there is some evidence to suggest that positive and negative valance (positive/negative) words and prosody may be processed differently (e.g., Schirmer, Kotz & Friederici, 2002). More specifically Schirmer and colleagues (2002) primed participants with semantically neutral German sentences spoken in either happy or sad intonations and found that when intervals between the prime and target stimuli were short, men (but not women) demonstrated different reaction times and elicited different N400 amplitudes in response to positive or negative stimuli. The inverse was found when the inter-stimulus interval was extended to 750ms, where a differential effect for positive and negative words was found in women, but not men. This research, however nicely highlights that under some circumstances positive and negative information is processed differently, thus gives rise to the possibility that so too may differently valanced social intentions; an area of potential value for future research.

One potential explanation could be that there was no need to evaluate those constructs in more detail as the early (P200) analysis had already confirmed the speakers' different social intention. In fact, given the lack of task demands, it may well be that listeners ignored subtle prosodic differences between control and anger at this point because they were not going to be affected by these differences. The P200 component has been linked to more involuntary processing (c.f. Paulmann & Kotz, 2008); once this evaluation is completed, listeners may not feel compelled to focus on these materials further. However, in instances where early analysis has led to potentially ambiguous evaluation (e.g., I understand the speaker is in a positive state, but I have yet to understand what they want from me/how it affects me), later processing stages may not be "ignored" even when task demands are lacking. This hypothesis receives some support from the observation that neutral prosody is not brought back into focus again, either. Thus, the lack of task in this experiment may have led to only shallow attention being given to the stimuli and thus not processed in depth. Some studies have shown that task demands can not only influence processing patterns (e.g., Plante, Creusere & Sabin, 2002), but also may potentially affect cognitive processing (e.g., Sandi, 2013). Although the primary conclusions of Sandi's (2013) study centred on stress, task demands were identified as a factor mediating cognitive functioning, such as memory and goal-directed behaviour. It therefore seems reasonable to assume that the task demands of our study may have led to this effect.

Limitations and directions for future research

While the current investigation has enriched our understanding about the similarities and differences when expressing motivations and emotions through vocal parameters, such an investigation does not come without its drawbacks. In the

following, some potential limitations for this set of studies are discussed and recommendations for future research are provided.

Study 1 set out to explore acoustic parameters associated with expressing emotions and motivations through voice. Results showed that each investigated state is conveyed with a unique configuration of vocal cues and reinforced the notion that, on an acoustic level, motivations and emotions are communicated differently. The independent expression of states was found to be reliant on a complex combination of vocal cues, along with generalisable differences in speaker modulations when conveying particular states (e.g., cold-anger was spoken slowest, and joy was expressed with the largest increase in average pitch). As such findings reinforced both the notion that emotional states are communicated discretely, and that motivations and emotions are expressed differently through prosody. However, despite testing portrayals from more than the traditional four speakers, speaker variability was still limited to fourteen speakers, thus may not contain the same degree of variability in cue modulations that a natural sample may contain. The exclusive use of voice professionals, despite being noted to produce high quality, more prototypical portrayals (e.g., Banse & Scherer, 1996; Paulmann et al., 2016), has also been questioned for yielding portrayals with some over emphasised vocal cues (e.g., see Scherer, 2003; Kriesman & Sidtis, 2011, for discussions). Because the motivational qualities of interest are likely to be less pronounced in an empirical setting in contrasts to real-life motivational situations (e.g., Zougkou, Weinstein & Paulmann, 2017), these over emphasised vocal cues may have rendered motivational portrayals more in line with how lay speakers would motivate others in real life settings. Still, the limited sample and empirical setting cannot be ignored when considering the potential variability in how these states are conveyed

prosodically. Therefore, future studies could consider getting materials from live interactions and from larger participant pools.

Study 2 built on the previously established acoustic profiles, this time accounting for portrayal quality. Acoustic profiles found in this study predominantly corroborated those established in our initial study, but more than that they indicated the listeners do not always base their judgements on the acoustic cues the speakers modulate when intending to convey a particular message through prosody. Specifically, differences in voice quality parameters seemed to be less pronounced in the recognised exemplars, suggesting that judge decisions were not as reliant on voice quality. With that being said, voice quality played an important role in distinguishing between states when other acoustic parameters were similar, thus reinforcing the earlier suggestion that communication of these states is achieved via a complex configuration of vocal cues. Furthermore, it was noted that some vocal cues were more pronounced in the recognised portrayals, raising the possibility that certain cues (in this case pitch variability, speech rate and intensity variability) may be more influential in the effective transmission of the intended message than other vocal cues. Interestingly, precisely these cues have been reported as influential on P200 salience detection (e.g., e.g., Picton, et al.,1977; Pantev, et al., 1996; Chang, et al., 2018), in turn supporting this possibility. However, due to unavoidable differences in speaker intensity, the auditory files in our study were normalised to a constant mean amplitude and as a result important acoustic information might have been lost. With it posited that judges may predominantly rely on intensity cues when differentiating between emotional states (e.g., Banse & Scherer, 1996) and arousal, arguably most closely linked with intensity cues, eliciting stronger P200 responses

(Paulmann, Bleichner & Kotz, 2013), it is recommended that future studies take all possible precautions to avoid normalising the intensity of materials

Furthermore, despite demonstrating that motivational and emotional tones of voice differ in their cue configurations, the present studies were not geared towards providing an answer to which vocal cues listeners based their judgements on. By highlighting differences in acoustic profiles across states for unscreened and well-recognised utterances, the present investigation suggests that some cues may be more informative for judge differentiation. However, to isolate the precise cues upon which listeners base their decisions and whether there is a hierarchy to vocal cues a systematic investigation in which vocal cues are varied and judge recognition is assessed is required. A study of this nature would also be highly informative in both, reinforcing established acoustic profiles for motivational and emotional prosody as well as potentially add further weight to the separation of these types of prosody.

Study 3 explored ERP correlates associated with emotional and motivational prosody; we were interested in investigating how prosody is processed implicitly, thus more closely mirroring every day evaluations of prosodic features (e.g., it is rare for listeners to actively categorize what they just heard). In fact, there is some evidence to suggest that passive listening tasks may elicit weaker responses than those exhibited during active tasks (e.g., e.g., Böcker, Bastiaansen, Vroomen, Brunia & Gelder, 1999). Moreover, there is a well-developed literature on the influences of task focus and demands on ERP components (e.g., e.g., Plante, Creusere, & Sabin, 2002; Wildgruber, et al., 2004; Kotz & Paulmann, 2008; Chen, Zhao, Jiang & Yang, 2011; Wickens & Perry, 2015). As such the final study in this investigation attempted to avoid these issues by promoting engaging participants in a passive listening task by making them believe that it was important to focus on the materials and not drift

off entirely, yet, not giving participants any task may not have provoked the desired engagement with the stimuli. Furthermore, in the real-world it is likely exceedingly rare for a listener to simply sit still and try to detect a probe, remember a word or perform a similarly trivial task whilst listening to prosodic communications, thus highlighting another benefit of including a suitably designed task. Future research can directly address the issue of task demands by employing a between-subject design that provides one group of listeners with an active task and another with a passive task. One such task could be to ask participants to indicate in which context they think they would most likely hear or use the utterance after each presentation. A task such as this may enhance the saliency of the stimuli to participants, potentially leading to deeper stimuli processing. Similarly, by asking participants to categorise stimuli based on probable context of use it is likely that all stimuli will undergo a similar evaluation process, thus remain comparable. However, care would need to be given to ensure that semantic content did not become the primary focus of this evaluation process and if used with pseudo-sentences would need to be adapted accordingly. Nonetheless, a reassessment of the processing time-course of emotional and motivational prosody using a suitable task would be highly insightful in identifying similarities and differences between these prosodic constructs.

Naturally, all studies that solely focus on prosody processing ignore that real-world vocal exchanges often provide other sources of information. For instance, when the exchange takes place face-to-face, facial cues and contextual information may assist in the conveyance and interpretation of the intended message. Even in cases where facial cues are not available (e.g., telephone calls) and the intended message is seemingly solely reliant on prosody, it seems reasonable to contend that a number of contextual and psychosocial factors are still assisting to determine the

effectiveness of the exchange. Contextual factors could be as simple as the position of the listener in relation to the intended message (i.e., are they in a position to do exactly what the speaker wants done?). More complexly, the speaker-listener relationship status, their familiarity with each other, and even the experiences of either party which may have led to particular semantics being associated with specific contexts (e.g., “Tell me what you mean by this” may remind them of a specific experience) may impact the effectiveness of transmission and inference of prosodic communications. Work by Uskul, Paulmann and Weick (2016) highlights that social power (i.e., the capacity to control and shape their own resources and outcomes as well as those of others) may moderate the recognition of emotional states conveyed through prosody. Whilst there is little room for contention surrounding whether or not emotional states and social intentions can be communicated and recognised through prosody alone, more investigations into the factors, similar to social power, that may moderate this effect is needed. In terms of motivational tone of voice and other social intentions, this could be a systematic inquiry into precisely how much contextual information enhances and hinders recognition of the vocal communication of these states. In a similar vein, speaker-listener relationship and familiarity could be tested through testing prosody recognition on sub-samples of groups of people with varying familiarity and relationships. Investigations of these types would assist in answering important questions such as. Does knowing how a person normally sounds enhance your ability to recognise differences in their tone of voice? Or, is the recognition of motivational tones of voice highly contingent on contextual and psychosocial factors?

Conclusions

The present investigation set out to assess the similarities and differences between motivational and emotional tones of voice in an attempt to disseminate these constructs in the literature. Despite differences in the precise modulations of some acoustic parameters compared with the previous literature, the present investigation convincingly indicates that emotions and motivations, particularly autonomy-support, control, cold-anger and joy, are expressed through tones of voice comprised of different vocal cue configurations. By identifying generalisable differences in speaker modulations and acoustic differences in utterances recognised as conveying their intended state, this research not only reinforces the notion that motivational and emotional prosody is acoustically different, but also highlights the complexity of these differences. More precisely, in line with previous assertions (e.g., Paulmann, 2015), it was shown that the effective transmission and subsequent inference of emotional and motivational tones of voice is dependent on a complex configuration of acoustic parameters (i.e., each parameter may be modulated in a different way and/or with varied strength), which listeners must pick up on in order to correctly infer the intended message. In the final study, results suggested even though they are acoustically distinctly different, emotional and motivational tones of voice are processed at a similar point in time in the brain yet with the possibility that partly differing neural structures (indicated by distribution differences) may be at play. Taken together, the findings of these studies suggest that these types of prosody are likely distinct from one another, especially in the tones of voice used to express and recognise them, but on a neurophysiological level they are processed in a similar way.

References

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18, 172–184.

Bachorowski, J.-A., & Owren, M.J. (2008). Vocal Expressions of Emotion. In M. Lewis, J. M. Haviland-Jones & L.F. Barrett (Eds.). *Handbook of Emotions*, 3rd Ed (pp. 196-210). New York: The Guilford press.

Banse, R., & Scherer, K.R. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology*, 70 (3), 614-636.

Batson, C. D., Shaw, L. L., & Oleson, K. C. (1992). Differentiating affect, mood, and emotion: Toward functionally based conceptual distinctions. In M. S. Clark (Ed.), *Review of personality and social psychology, No. 13. Emotion* (pp. 294-326). Thousand Oaks, CA, US: Sage Publications, Inc.

Bezooijien, R. van. (1984). *The characteristics and recognizability of vocal expression of emotions*. Drodrecht, The Netherlands: Foris.

Besson, M., Magne, C., Schön, D., 2002. Emotional prosody: sex differences in sensitivity to speech melody. Trends in Cognitice science. 6, 405–407.

Blanc, J.M., & Dominey, P.F. (2003). Identification of prosodic attitudes by temporal recurrent network. *Cognitive Brain Research*, 17, 693-699.

Böcker, K. B., Bastiaansen, M., Vroomen, J., Brunia, C. H., & Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology*, *36*(6), 706-720.

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer*. Software.

Borden, G.J., & Harris, K.S. (1984). *Speech Science primer: Physiology, acoustics and perceptions of speech*. Baltimore: Williams & Wilkins.

Bosantov, V. & Kotchoubey, B. (2004). Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations. *Psychophysiology*, 41 (2), 259-268.

Bromberg-Martin, E.S, Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational control: rewarding, adversive, and alerting. *Neuron*, 68 (5), 815-834.

Castro, S.L., & Lima, C.F. (2010). Recognizing emotions in spoken language: A validated set of Portuguese sentences and pseudosentences for research on emotional prosody. *Behavior Research Methods*, 42 (1), 74-81.

Chang, J., Zhang, X., Zhang, Q., & Sun, Y. (2018). Investigating duration effects of emotional speech using stimuli in a tonal language by using event-related potentials. *IEE Access*, 6, 1-9.

Cheang, H.S., and Pell, M.D. (2008). The Sound of Sarcasm. *Speech Communication*, 50, 366-381.

Chen, X, Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of expectancy violation during emotional prosody processing. *Biological Psychology*, 86, 158-167.

Deci, E.L., & Ryan, R.M. (1987). The support of autonomy and the control of behavior. *Journal of Personality and Social Psychology*, 53 (6), 1024-1037.

Deci, E.L., & Ryan, R.M. (2000). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development and Well-Being. *American Psychologist*, 55 (1), 68-78.

Eckstein, K., & Friederici, A. D. (2006). It's early: event-related potential evidence for initial interaction of syntax and prosody in speech comprehension. *Journal of Cognitive Neuroscience*, *18*(10), 1696-1711.

Elowsson & Friberg (2017). Long-term Average Spectrum in Popular Music and its Relation to the Level of the Percussion. Proceedings of the 142nd Audio Engineering Society Convention.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.

Flippi, P., Congdon, J.V., Hoang, J., Bowling, D.L, Reber, S.A., Pasukonis, A., Hoeschele, M., Ocklenburg, S., de Boer, B., Sturdy, C.B., Newen, A., & Gunturkun, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings for Royal Society of Biology*, 284.

Fontaine, J.J.R., & Scherer, K.R. (2013). Emotion is for doing: the action tendency component. In: J.J.R. Fontaine, K.R. Scherer, C. Soriano (Eds.). *Components of emotional meaning; a sourcebook.* Oxford: Oxford University Press. pp.170–185

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. (2007). The world of emotion is not two-dimensional. *Psychological Science*, 13, 1050-1057

Geiser, E., Zaehle, T., Jancke, L., & Meyer, M. (2008). The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*, *20*(3), 541-552.

Grichkovtsova, I., Morel, M., & Lacheret, A. (2012). The role of voice quality and prosodic contour in affective speech perception. *Speech Communication*, 54, 414-429.

Hillyard, S.A., Hink, R.F., Schwent, V.L., & Picton T.W. (1973). Electrical signs of selective attention in the Human Brain. *Science*, 182, 177-180.

Inquisit 5 [Computer Software]. (2016). Retrieved from: https://www.millisecond.com/

Iredale, J.M., Rushby, J.A., McDonald, S., Dimoska-Di Marco, A. Swift, J. (2013). Emotion in voice matters: Neural correlates of emotional prosody perception. *International Journal of Psychophysiology*, 89 (3), 483-490.

Isen, A.M., Reeve, J. (2006). The influence of positive affect on intrinsic and extrinsic motivation: Facilitating enjoyment of play, responsible work behavior, and self-control. *Motivation and Emotion*, 29, 297–325.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Englewood Cliffs, NJ: Prentice-Hall.

Kotz, S.A., & Paulmann, S. (2007). When emotional prosody and semantics dance cheek to cheek: ERP evidence. *Brain Research*, 1151, 107-118.

Kotz, S.A., & Paulmann, S. (2011). Emotion, Language, and the Brain. *Language and Linguistics Compass*, 5 (3), 108.125.

Kraus, M.W. (2017). Voice-Only Communication Enhances Empathic Accuracy. *American Psychologist*, 72 (2), 644-654.

Kreiman & Sidtis (2013). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Oxford: Wiley.

Lima, C.F., & Castro, S.L. (2011). Speaking to the Trained Ear: Musical Expertise Enhances the Recognition of Emotions in Speech Prosody. *Emotion*, 11 (5), 1021-1031.

Levenson, R.W., & Ruef, A.M. (1992). Empathy: A Physiological substrate. *Journal of Personality and Social Psychology*, 63, 234-246.

Liu, T., Pinheiro, A. P., Deng, G., Nestor, P. G., McCarley, R. W., & Niznikiewicz, M. A. (2012). Electrophysiological insights into processing nonverbal emotional vocalizations. *NeuroReport*, *23*(2), 108-112.

Matsui, T., Nakamura, T., Utsumi, A., Saski, A.T., Koike, T., Yoshida, Y., Harada, T., Tanabe, H.C., & Sadato, N. (2016). The role of prosody and context in sarcasm comprehension: Behavioural and fMRI evidence. *Neuropsychologia*, 87, 74-84.

Matsumoto, D. (2002). Methodological requirements to test a possible In-Group Advantage in Judging Emotions Across Cultures: Comment on Elfenbein and Ambady (2002) and Evidence. *Psychological Bulletin*, 128 (2), 236-242.

Matsumoto, D. (2007). Apples and oranges: Methodological requirements for testing a possible ingroup advantage in emotion judgments from facial expressions. In Hess, U., and Philippot, P. (eds.), *Group dynamics and emotional expression* (pp. 140-181). New York: Cambridge University Press.

Matsumoto, D., & Eckman, P. (1989). American-Japanese cultural differences in intensity ratings of facial expressions of emotion. *Motivation and Emotion*, 13, 143-157.

Matsumoto, D., Consolacion, T., Yamanda, H., Suzuki, R., Franklin, B., Paul, S., Ray, R., & Uchida, H. (2002). American-Japanese cultural differences in judgements of emotional expression of different intensities. *Cognition and Emotion*, 16 (6), 721-747.

Menn, L. & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, 25 (4), 341-383.

Meyer, D. K., & Turner, J.C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18, 377–390.

Mitchel, R.L.C., & Ross, E. (2013). Attitudinal Prosody: What we know and Directions for future research. *Neuroscience and Behavioral Reviews, 37 (3)*, 471-479.

Mittal, R., Erath, B.D., & Pleniak, M.W. (2013) Fluid Dynamics of Human Phonation and Speech. *Annual Review of Fluid Mechanics*, 45, 437-467.

Monson, B.B., Hunter, E.J., Lotto, A.J., & Story, B.H. (2014). The perceptual significance of high-frequency energy in the human voice. *Frontiers in Psychology*, 5, 1-11.

Mozziconacci, S.J.L. (2001). Modelling emotion and attitude in speech by means of perceptually based parameter values. *User Modelling and User-Adapted Interaction*, 11, 297-326.

Murray, I.R., & Arnott, J.L. (1993). Toward the stimulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of Acoustical Society of America*, 93 (2), 1097-1108.

Niemiec, C.P., & Ryan, R.M. (2009). Autonomy, competence, and relatedness in the classroom. Applying self-determination theory to educational practice. *Theory and Research in Education*, 7 (2), 133-144.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9 (1), 97-113.

Owren, M.J., & Bachorowski, J.-A. (2007). Measuring vocal acoustics. In J.A. Coan & J.J.B. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 239-266). New York: Cambridge University Press.

Owren, M.J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate signaling. In D.H. Owings, M.D. Beecher, & N.S. Thompson (Eds.). Perspectives in ethology; Vol. 12. Communication (pp. 299-346). New York: Plenum Press.

Owren, M.J., Rendall, D., & Bachorowski, J.-A. (2003). Nonlinguistic vocal communication. In D. Maestripieri (Ed.). Primate Psychology (pp. 359-394). Cambridge MA: Harvard University Press.

Pantev, C., Elbert, T., Ross, B., Eulitz, C. & Terhardt, E. (1996). Binaural fusion and the representation of virtual pitch in the human auditory cortex. *Hearing Research*, 100 (1), 164-170.

Paulmann, S. (2015). The Neurocognition of Prosody. In G. Hickok and S. Small (Eds.). *Neurobiology of Language* (pp. 1109-1118). San Diego: Elsevier.

Paulmann, S., Bleichner, M. & Kotz, S.A. (2013). Valence, arousal and task effects in emotional prosody processing. *Frontiers in Psychology*, 4, 345.

Paulmann, S., Furnes, D., Boknes, A.M., & Cozzolino, P.J. (2016). How Psychological Stress Affects Emotional Prosody. *Plos One*, 11 (11).

Paulmann, S. & Pell, M.D., & Kotz, S.A. (2010). How aging affects the recognition of emotional speech. *Brain and Language*, 104, 262-269.

Paulmann, S., Jessen, S., & Kotz, S.A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia*, 50, 1609-1620.

Paulmann, S. & Kotz, S.A. (2008). Early emotional prosody perception based on different speaker voices. *Neuroreport*, 19 (2), 209-213.

Paulmann, S., & Kotz, S.A. (In Press). The electrophysiology and time-course of vocal emotion expressions. In S. Fruehholz & P. Belin (Eds.) *Oxford Handbook of Voice Perception* (Chapter 20)

Paulmann, S., Ott, D.V.M., Kotz, S.A. (2011), Emotional Speech perception unfolding in time: The role of the basal ganglia. *Plos One*, 6 (3), 1451-1454.

Paulmann, S., & Uskul, A.K. (2014). Cross-cultural emotional prosody recognition: Evidence from Chinese and British Listeners. *Cognition and Emotion*, 28 (2), 230-244.

Paulmann, S., Vrijders, B., Weinstein, NB., & Vansteenkiste, M. (2018). How parents motivate their children through prosody. Proceedings of the 9th International Conference on Speech Prosody.

Pell, M.D. (2006). Cerebral mechanisms for understanding emotional prosody in Speech. *Brain and Language*, 96 (2), 221-234.

Pell, M.D., Monetta, L., Paulmann, S., & Kotz, S.A. (2009). Recognising emotions in a foreign language. *Journal of Nonverbal Behavior*, 33, 107-120.

Pell, M.D., Paulmann, S., Dara, C., Alasseri, A., & Kotz, S.A. (2009). Factors in the recognition of vocally expressed emotions: A comparison of four languages. *Journal of Phonetics*, 37, 417-435.

Pell, M.D., Rothermich, K., Liu, P, Paulmann, S., Sethi, S., Rigoulot, S.(2015). Preferential decoding of emotion from human non-linguistic vocalizations versus speech prosody. *Biological, Psychology,* 111, *14-25.*

Picton, T. W., & Hillyard, S. A. (1988). Endogenous event-related potentials. In: T. W. Picton (Ed.), *Handbook of electroencephalography and clinical neurophysiology* (Vol. 3, pp. 361–427). Amsterdam: Elsevier.

Picton, T.W., Woods, D.L., Baribeau-Braun, J. & Healey, T.M. (1977). Evoked potential audiometry. Journal of Otolaryngology, 6 (2), 90-119.

Pittam, J., & Scherer, K.R. (1993). Vocal expression and communication of emotion. In M. Lewis & J.M. Haviland (Eds.), *Handbook of Emotions* (pp. 185-198). New York: Guilford Press.

Planalp, S. (1998). Communicating emotion in everyday life: cues, channels, and processes. In A. Andersen and L.K. Guerrero (Eds.), *Handbook of communication and emotion* (pp. 29–48). New York: Academic Press.

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating sentential prosody from sentence processing:activation interacts with task demands. *NeuroImage*, *17*(1), 401-410.

Rabinov, C.R., Kreiman, J., Gerratt, B.R., & Bielamowicz, S. (1995). Comparing Reliability of Perceptual Ratings of Roughness and Acoustic Measures of Jitter. *Journal of Speech, Language, and Hearing Research*, 38, 26-32.

Radel, R., Sarrazin, P., and Pelletier, L. (2009). Evidence of subliminally primed motivational orientations: the effects of unconscious motivational processes on the performance of a new motor task. *Journal of Sport and Exercise Psychology*, 31 (5), 657-674.

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, *44*(3), 159-175.

Regel, S., Gunter, T.C., Friederici, A.D. (2010). Isn't it Ironic? An Electrophysiological Exploration of Figurative Language Processing. *Journal of Cognitive Neuroscience*, 23 (2), 277-293.

Rigoulot, S., Fish, K., & Pell, M. (2014). Neural correlates of inferring speaker sincerity from white lies: an event-related potential source localization study. *Brain Research*, 1656, 48-62.

Russell, J.A., Bachorowski, J.-A., & Fernandez-Dols, J.-M. (2003). Facial and vocal expressions of emotion. Annual Review of Psychology, 54, 329-349.

Ryan, R.M., & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55 (1), 68-78.

Sandi, C. (2013). Stress and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4, 245-261.

Scherer, K. R. (1979). Non-linguistic indicators of emotion and psychopathology. In C. E. Izard (Ed.), *Emotions in personality and psychopathology* (pp. 495-529). New York: Plenum.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K.R. Scherer & P. Ekman (Eds.). *Approaches to Emotion* (pp. 293-318). Hillsdale, NJ: Erlbaum.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143-165.

Scherer, K.R. (2003). Vocal communication of Emotion: A Review of Research Paradigms. *Speech communication*, 40, 227-256.

Scherer KR (2004). Feelings integrate the central representation of appraisal-driven response organization in emotion. In: A. SR Manstead, NH Frijda, AH Fischer (Eds.), *Feelings and emotions: the Amsterdam symposium* (pp.136–157). Cambridge, Cambridge University Press.

Scherer, K. R., Banse, R., Wallbott, H. G., & Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion*, 15, 123-148.

Schirmer, A., Chen, C., Ching, A., Tan, L., Hong R.Y. (2013). Vocal emotions influence verbal memory: Neural correlates and interindividual differences. *Cognitive, Affective & Behavioural Neuroscience*, 13 (1), 80-93.

Schirmer, A., Kotz, S.A., Friederici, A.D., 2002. Sex differentiates the role of emotional prosody during word processing. Cognition and Brain Research, 14, 228–233.

Schirmer, A., Kotz, S.A., Friederici, A.D., 2005. On the role of attention for the processing of emotions in speech: sex differences revisited. *Cognitive Brain Research*, 24 (3), 442–452.

Schirmer, A. & Kotz, S.A. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences*, 10, 24-30.

Sobin, C., & Alpert, M. (1999). Emotion in Speech: The Acoustic Attributes of Fear, Anger, Sadness, and Joy. *Journal of Psycholinguistic Research*, 28 (4), 347-365.

Soenens, B., & Vansteenkiste, M. (2010). A theoretical upgrade of the concept of parental psychological control: Proposing new insights on the basis of self-determination theory. *Developmental Review*, 30, 74-99.

Steinhauer, K., Alter, K., & Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191–196.

Stekelenburg, J.J. & Vroomen, J.J.S. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Frontiers in Integrated Neuroscience*, 6 (6), 25

Stevens, K.N. (2000). *Acoustic Phonetics*. Cambridge MA: MIT Press.

SuperLab 5 [Computer Software] (2015). Retrieved from: https://www.cedrus.com/superlab/

Uskul, A.K., Paulmann, S., & Weick, M. (2016). Social Power and Recognition of Emotion Prosody: High Power Is Associated with Lower Recognition Accuracy Than Low Power. *Emotion*, 16 (1), 11-15.

Vandercammen, L., Hofmans, J., & Theuns, P. (2014). The mediating role of affect in the relationship between need satisfaction and autonomous motivation. *Journal of Occupational and Organizational Psychology*, 87, 62–79.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., Parks, M., (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 394–417.

Van Petten, C., Kutas, M., 1988. The use of event-related potentials in the study of brain asymmetries. *International Journal of Neuroscience*, 39, 91–99.

Van Petten, C., Luka, B.J., (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. Brain and Language. 97(3), 279–293.

Wambacq, I. J., & Jerger, J. F. (2004). Processing of affective prosody and lexical-semantics in spoken utterances as differentiated by event-related potentials. *Cognitive Brain Research*, 20(3), 427-437.

Wickens, S., & Perry, C. (2015). What Do You Mean by That?! An Electrophysiological Study of Emotional and Attitudinal Prosody. *Plos One*, 10(7)

Weinstein, N., & Hodgins, H.S. (2009). The moderating role of autonomy and control on the benefits of written emotion expression. *Personality and Social Psychology Bulletin*, 35 (3), 351-364.

Weinstein, N., & Ryan, R.M. (2010). When Helping Helps: Autonomous Motivation for Prosocial Behavior and Its Influence on Well-Being for the Helper and Recipient. *Journal of Personality and Social Psychology*, 98 (2), 222-244.

Weinstein, N., Zougkou, K., Paulmann, S. (2014). Differences between the acoustic typology of autonomy-supportive and controlling sentences. In: *Proceedings of the 7th Conference on Speech Prosody 2014*, Dublin, Ireland.

Weinstein, N., Zougkou, K., & Paulmann, S (2018). You 'have' to hear this. Using tone of voice to motivate others. *Journal of Experimental Psychology: Human Perception and Performance*, 44 (6), 898-913.

Wickens, S., & Perry, C. (2015). What Do You Mean by That?! An Electrophysiological Study of Emotion and Attitudinal Prosody. *Plos One*, 10 (7).

Wildegruber, D., Hertrich, I., Riecker, A., Erb, M., Anders, S., Grodd, W. & Ackermann, H. (2004). Distinct frontal regions subserve evaluation of linguistic and emotional aspects of speech intonation. *Cerebral cortex*, 14 (12), 1384-1389.

Williams, C.E., & Stevens, K. N. (1972). Emotions and Speech: Some Acoustic correlates. *Journal of the Acoustical Society of America*, 1238-1250.

Wolfe, V., Fitch, J. & Cornell, R. (1994). Acoustic Prediction of Severity in Commonly Occurring Voice Problems. *Journal of Speech, Language, and Hearing Research*, 38, 273-279.

Yumoto, E., Gould, W.J., & Baer, T. (1982). Harmonics-to-noise ration as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 71 (6), 1544-1550.

Zougkou, K., Weinstein, N., & Paulmann, S. (2017). ERP correlates of Motivating Voices: Quality of Motivation and Time-Course Matters. *Social Cognitive and Affective Neuroscience*, 12 (10), 1687-1700.

Appendices

Appendix 1: Sentence list

ID	**Sentence**	**ID**	**Sentence**
1	Which one do you recommend?	36	When will you be free?
2	It's up to you to complete	37	It's ready to go
3	This calls for you to focus	38	I'm free if you want to talk.
4	Will you come visit me?	39	why don't you ask for help?
5	When are you planning to start?	40	Can you give me a hand with this?
6	you can stay in here	41	You can meet me tomorrow
7	You can't keep this up	42	I suggest you go with the alternative
8	Why don’t you go there?	43	Bring that over here
9	Come to visit me	44	Take a look at this
10	Keep an eye out for it	45	You can meet me there
11	Is this how you want it to be?	46	Tell me more about this
12	Can you keep on trying?	47	Have you considered changing something?
13	It's time to leave	48	I'll wait for your call
14	How long before you finish?	49	Can you finish this?
15	Do you want to do this?	50	Why don't you take a break?
16	Can you help me do this?	51	Why don’t we try tomorrow
17	Any way you can speed this up?	52	I suggest you reconsider that
18	Someone needs to do it.	53	I recommend you pay attention
19	Can you bring that over here?	54	I suggest you get help with this
20	Is this what you had in mind?	55	I'll wait for you to call me
21	When will you have it finished?	56	Consider if you want to continue like this
22	Have you read this?	57	Consider changing something here
23	Can you stay in here?	58	Show me what you mean
24	Do you want help with something?	59	Giving me more information will help
25	When will you have it done?	60	Show me how you did it
26	Can you do it like this?	61	Decide how you want to do this
27	When do you expect to finish it?	62	Is this what you want to do?
28	Do you know how to do this?	63	Can you check this?
29	You can use this here	64	I suggest you don't rush this
30	Why don’t you finish it!	65	I suggest that you consider this
31	Why don’t you go ahead	66	I recommend you give it some thought
32	It’s time to go	67	Call me once you're done
33	wait there for me	68	Tell me when you are ready
34	Keep on going like this	69	Tell me what you mean by this.
35	Is this something you can do?	70	let me know me when you finish it.

Appendix 2: Study 1 demographics and descriptives

Encoders

Encoder	Age	Sex	Acting years	Accent
1	20	F	5	Manchester
2	18	M	3	Suffolk
3	21	F	3	London
4	19	M	8	South UK
5	20	F	5	Midlands
6	22	F	7	London
7	18	M	10	East Anglia
8	19	F	12	Yorkshire
9	18	M	3	Milton Keynes
10	24	M	10	Essex
11 (Not used)	N/A	N/A	N/A	U.S.A
12	18	F	3	Midlands
13	20	M	14	Herts.
14	23	F	3	Southwest
15	21	M	14	Norfolk

Demographic summary

	N	Minimum	Maximum	Mean	Std. Deviation
Age	14	18	24	20.07	1.940
Years	14	3	14	7.14	4.204
Valid N (listwise)	14				

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	7	50.0	50.0	50.0
	Female	7	50.0	50.0	100.0
	Total	14	100.0	100.0	

Acoustics Descriptives of unscreened sample

Anger

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	948	.003818760560	1.844545463000	.41966999600000	.270369221000000
NrangeF0	948	.166863762000	4.506142468000	1.30062685300000	1.043828772000000
rangedB	948	17.56093228	71.96565198	37.0941800000	9.04045602600
Sprate	948	.088974553	.539824263	.21115419100	.056131275000
EB_0_500	948	30.20380841	51.96448549	39.6799351400	2.65135896700
EB_0_1000	948	30.60074548	49.21100388	38.6265666500	1.98529961300

EB_500_1000	948	14.30142603	48.08433270	35.3759584800	4.40136017600
EB_1000_5000	948	12.900669150	38.077175770	24.95939523000	4.171956370000
EB__5000_8000	948	-2.884030865	29.024489000	11.27411552000	4.996129522000
Valid N (listwise)	948				

Autonomy

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	959	.072776705000	2.013922003000	.58226725700000	.318966950000000
NrangeF0	959	.147467799000	4.201220992000	1.28786205300000	.844779790000000
rangedB	959	16.02558584	69.75407415	34.7246416400	8.04704762700
Sprate	959	.087679516	.418737717	.19127392500	.047238706900
EB_0_500	959	32.21818255	49.50986611	40.0864375300	2.19730143900
EB_0_1000	959	31.99742741	49.12989659	38.8463290700	1.69831550300
EB_500_1000	959	18.61122464	48.71346773	35.3054055000	4.24675688600
EB_1000_5000	959	13.207657170	33.988628020	24.36012886000	3.603335237000
EB__5000_8000	959	-1.699323767	29.243961550	12.01620742000	4.979062682000
Valid N (listwise)	959				

Control

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	966	.018482585500	1.225641576000	.45160452200000	.212657470000000
NrangeF0	966	.132099711000	4.307500466000	1.15502907500000	.831138967000000
rangedB	966	17.53135906	67.84542162	36.3084635600	8.70770370200
Sprate	966	.086581003	.476000000	.19645474500	.051515395300
EB_0_500	966	30.98163096	50.00910041	39.7334475300	2.70717442000
EB_0_1000	966	32.71364654	48.39467246	38.7926935800	1.92415048900
EB_500_1000	966	16.23494256	47.23652998	35.9158805800	3.94866302700
EB_1000_5000	966	9.850344984	37.692309970	25.35936500000	3.774998772000
EB__5000_8000	966	-3.393426960	30.568759790	11.79798749000	5.316661298000
Valid N (listwise)	966				

Joy

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	958	.115496594000	2.075756022000	.84991269200000	.395074461000000
NrangeF0	958	.243335624000	4.198566109000	1.52537150800000	.785367804000000
rangedB	958	15.45922870	69.86851046	34.7619063500	8.32868705700
Sprate	958	.078332074	.345073696	.17508254800	.039763898900
EB_0_500	958	29.65620535	53.30747531	39.2128503100	2.83506797300
EB_0_1000	958	32.11715776	50.40448928	38.4804752200	1.93293194000
EB_500_1000	958	13.21209984	49.33379943	35.7833759300	4.02122011400
EB_1000_5000	958	8.869960044	39.119558370	25.58518571000	3.676079645000
EB__5000_8000	958	-5.133000057	27.041054460	12.64445010000	5.658352702000
Valid N (listwise)	958				

Neutral

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	964	.029437061000	2.169568244000	.30265621400000	.191761071000000
NrangeF0	964	.117975153000	4.516025711000	1.08892366700000	1.033332107000000
rangedB	964	13.31561874	67.23519752	34.4519791400	7.36013780700
Sprate	964	.067023294	.384123745	.19393002300	.042009924900
EB_0_500	964	30.86178831	53.25293821	40.2974663500	2.58176546600
EB_0_1000	964	33.26884020	50.78261306	39.1306904100	1.68294982900
EB_500_1000	964	20.03195205	45.96322827	35.8005304900	3.62927449700
EB_1000_5000	964	10.687264290	35.064572230	24.12727301000	3.934023519000
EB__5000_8000	964	-2.676948988	32.757779720	10.86983965000	5.786657971000
Valid N (listwise)	964				

Appendix 3: Study 2 demographics and descriptives

Demographics of stimuli validation study

Vision

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	306	81.0	81.0	81.0
	Yes	72	19.0	19.0	100.0
	Total	378	100.0	100.0	

Neurological

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	363	96.0	96.0	96.0
	Yes	15	4.0	4.0	100.0
	Total	378	100.0	100.0	

Mental

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	343	90.7	90.7	90.7
	Yes	35	9.3	9.3	100.0
	Total	378	100.0	100.0	

Hearing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	378	100.0	100.0	100.0

Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	287	75.9	75.9	75.9
	Male	91	24.1	24.1	100.0
	Total	378	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Education	378	6	20	13.74	2.881
Age	378	16	55	20.03	3.524
Valid N (listwise)	378				

Acoustics Descriptives of recognised files

Anger

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	194	.003818760560	1.844545463000	.42347760900000	.26686396700000
NrangeF0	194	.181039147000	4.340114887000	1.631565956000	1.294428330000
EB_0_500	194	31.11618497	48.35908429	39.8939236600	2.43619911100
EB_0_1000	194	31.62479748	45.64439841	38.6470534800	1.80809112200
EB_500_1000	194	14.30142603	42.93888087	34.8097970700	4.71302932700
EB_1000_5000	194	14.946521800	32.685520790	25.04662717000	3.950090626000
EB__5000_8000	194	-.644038131	29.024489000	11.44487831000	5.133722455000
Sprate	194	.121093474	.539824263	.24591723100	.070119201700
meandB	194	56.09908377	68.22998685	64.0109445100	2.35783597600
rangedB	194	18.71605470	67.04606319	42.0693149700	9.27545295200
Valid N (listwise)	194				

Autonomy

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	292	.092389102300	1.749511071000	.60585191200000	.29602071800000
NrangeF0	292	.316068113000	4.201220992000	1.318883771000	.71980599000000
EB_0_500	292	32.21818255	49.50986611	39.9366890000	2.30948329600
EB_0_1000	292	31.99742741	49.12989659	38.7941117700	1.78569017500
EB_500_1000	292	21.43757733	48.71346773	35.3712692200	4.33630399800
EB_1000_5000	292	13.496937580	33.727791620	24.00261926000	3.972451042000
EB__5000_8000	292	-1.544246849	29.243961550	11.90593953000	5.554951585000
Sprate	292	.087679516	.418737717	.18960309600	.045723176900
meandB	292	57.76430963	74.12938816	65.8893091100	1.91244826800
rangedB	292	17.91047103	62.60680954	35.4776626200	8.05293810500
Valid N (listwise)	292				

Control

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	200	.080176904700	1.053537653000	.43628417600000	.180865447000000
NrangeF0	200	.251986180000	4.250947191000	1.14634668100000	.740876802000000
EB_0_500	200	32.18217874	45.13929302	39.3730713400	2.57116768000
EB_0_1000	200	33.47983857	46.31329642	38.6522724400	1.73000769700
EB_500_1000	200	22.42094495	47.23652998	36.2764119500	3.75203990000
EB_1000_5000	200	9.850344984	32.844930160	24.78986778000	3.854170262000
EB__5000_8000	200	-3.393426960	25.456413500	12.04401520000	5.118404578000
Sprate	200	.110899471	.383941799	.21973006600	.057658288100
meandB	200	53.78720231	72.04229924	65.0530770400	2.15556382900
rangedB	200	20.52634495	62.51045382	37.8418387700	8.43040162500
Valid N (listwise)	200				

Joy

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	237	.185827874000	2.001605621000	.93049754100000	.391334748000000
NrangeF0	237	.307810612000	3.799523700000	1.54218897800000	.685477551000000
EB_0_500	237	30.24421995	45.04850648	39.2596950400	2.72822539000
EB_0_1000	237	33.83366756	47.69955806	38.6008548400	1.60935715900
EB_500_1000	237	24.44229884	49.33379943	36.0814264100	3.63636267700
EB_1000_5000	237	8.869960044	37.236880000	25.05906908000	3.991943583000
EB__5000_8000	237	-5.133000057	24.836701440	9.67338791500	6.417445640000
Sprate	237	.078332074	.293905896	.18104242000	.035124238200
meandB	237	59.53612588	70.95225100	65.4430884400	1.83858044400
rangedB	237	17.36767165	54.36879244	34.3378960200	7.73254025900
Valid N (listwise)	237				

Neutral

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	817	.029437061000	2.169568244000	.29175973200000	.193946015000000
NrangeF0	817	.120205168000	4.459815690000	1.09375315300000	1.060622938000000
EB_0_500	817	30.86178831	53.25293821	40.5222215000	2.39414072800
EB_0_1000	817	33.26884020	50.78261306	39.1585246200	1.63662734000
EB_500_1000	817	20.03195205	45.31919291	35.5140947500	3.58301020000
EB_1000_5000	817	12.223325260	35.064572230	23.72765908000	3.804209285000
EB__5000_8000	817	-2.676948988	32.757779720	10.53448092000	5.508008623000
Sprate	817	.067023294	.384123745	.19375820700	.042300890800
meandB	817	59.58675930	76.03511819	65.8268080900	1.64601080500
rangedB	817	13.31561874	67.23519752	34.0150848300	7.10608083000
Valid N (listwise)	817				

Appendix 4: Study 3 demographics and descriptives

Descriptives for EEG recognition study

Participant	Age	Gender	Education	Hearing	Language	Mental	Neurological	Vision
1	22	2	16	1	English	1	1	1
2	18	1	8	1	English	1	1	1
3	20	2	6	1	English	2	1	1
4	20	2	15	1	English	1	1	1
5	19	2	14	1	English	1	1	1
6	20	2	12	1	English	1	1	1
7	20	2	16	1	English	1	1	1
8	18	2	7	1	English	1	1	1
9	25	1	15	1	English	1	1	1
10	20	1	13	1	English	1	1	1
11	50	1	15	1	English	1	1	1
12	53	2	17	1	English	1	1	1
13	21	2	17	1	English	2	1	1
14	21	2	16	1	English	1	1	2
15	23	2	16	1	English	1	1	1
16	22	2	16	1	English	1	1	1
17	43	2	15	1	English	1	1	1
18	25	2	18	1	English	1	1	1
19	21	2	14	1	English	1	1	1
20	54	2	18	1	English	1	1	1
21	27	1	12	1	English	1	1	1
22	50	2	22	1	English	1	1	1
23	24	2	18	1	English	1	1	2
24	18	2	10	1	English	1	1	1
25	19	2	16	1	English	1	1	1
26	18	2	14	1	English	1	1	1
27	19	2	14	1	English	1	1	2
28	19	2	10	1	English	1	1	1
29	19	2	16	1	English	1	1	2
30	19	2	14	1	English	1	1	1
31	20	2	6	1	English	2	1	1
32	20	2	16	1	English	1	1	1
33	21	2	14	1	English	1	1	2
34	19	2	14	1	English	1	1	1
35	19	2	14	1	English	1	1	1
36	20	1	13	1	English	1	1	1
37	20	2	15	1	English	1	1	2
38	19	2	15	1	English	2	1	1
39	19	2	15	1	English	1	1	1
40	19	1	13	1	English	1	2	1
41	23	1	19	1	English	1	2	2

42	33	1	12	1	English	1	1	1
43	48	2	15	1	English	1	1	1
44	55	2	10	1	English	1	1	1
45	43	1	15	1	English	1	1	1
46	39	1	15	1	English	2	1	1
47	51	1	15	1	English	1	1	1
48	34	1	15	1	English	1	1	2
49	37	1	13	1	English	1	1	1
50	27	2	18	1	English	2	1	1

Statistics

		Sex	Hearing	Mental	Language	Neurological	Vision
N	Valid	50	50	50	50	50	50
	Missing	0	0	0	0	0	0

Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	14	28.0	28.0	28.0
	Female	36	72.0	72.0	100.0
	Total	50	100.0	100.0	

Hearing

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	50	100.0	100.0	100.0

Mental

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	44	88.0	88.0	88.0
	Yes	6	12.0	12.0	100.0
	Total	50	100.0	100.0	

Language

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	English	50	100.0	100.0	100.0

Neurological

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	48	96.0	96.0	96.0
	Yes	2	4.0	4.0	100.0
	Total	50	100.0	100.0	

Vision

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	42	84.0	84.0	84.0
	Yes	8	16.0	16.0	100.0
	Total	50	100.0	100.0	

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Education	50	6	22	14.24	3.159
Age	50	18	55	27.06	11.795
Valid N (listwise)	50				

Demographics for EEG

PPt	Gender	Age	Sleep	Hours	Conc	ConcChan	EEG
1	F	19	O	4	O	Changing	Y
2	F	19	G	8	G	Changing	N
3	F	19	B	2.5	O	Changing	N
4	F	23	G	8	G	Worse	Y
5	F	19	O	7	G	Changing	Y
6	M	19	O	8	G	Better	N
7	M	19	G	9	G	Changing	N
8	N/A	N/A	N/A	N/A	N/A	N/A	N/A
9	N/A	N/A	N/A	N/A	N/A	N/A	N/A
10	M	19	O	6	G	Worse	N
11	F	23	O	5	G	Changing	Y
12	F	20	O	7	G	Worse	Y
13	F	20	O	9	O	Worse	N
14	M	21	G	10	G	Worse	Y
15	F	20	O	6	G	Changing	N
16	M	19	O	5	G	Better	Y
17	F	22	G	8	G	Worse	Y
18	F	22	G	7	G	Changing	Y
19	M	19	G	10	O	Worse	N
20	F	22	G	7	O	Same	N
21	M	19	G	9.5	G	Changing	N
22	M	19	G	8.5	O	Better	N
23	F	22	G	9	G	Worse	N
24	F	19	G	7.5	G	Same	Y
25	F	19	G	8	O	Better	N
26	M	23	G	8	G	Same	Y
27	M	21	G	6	O	Better	N
28	M	21	O	7	O	Worse	N
29	M	21	G	6	G	Changing	Y
30	M	22	B	0	O	Changing	Y

31	M	20	G	8	G	Better	N
32	M	19	G	7	O	Worse	N
33	F	21	O	7	O	Same	N
34	M	18	O	7	O	Changing	N
35	M	24	G	7.5	G	Changing	Y
36	M	20	G	7	G	Worse	N
37	N/A	N/A	N/A	N/A	N/A	N/A	N/A
38	F	26	G	8	G	Changing	Y
39	M	27	B	6	G	Better	N
40	M	32	O	7	G	Same	N
41	M	22	G	7	O	Better	N

Questionnaire Smoke & Drink:

PPt	Smoke	Amount	Last	Drink	Amount	Last
1	N			Y	Monthly	2 Weeks
2	Y	Weekly	Yeterday	Y	Weekly	Yesterday
3	N			Y	Weekly	2 Weeks
4	Y	2/Day	Today	Y	Monthly	Month
5	Y	5/day	Today	Y	Monthly	Yesterday
6	N			Y	B-Weekly	Week
7	N			N		
8	N/A	N/A	N/A	N/A	N/A	N/A
9	N/A	N/A	N/A	N/A	N/A	N/A
10	N			N		
11	N			Y	Weekly	Week
12	N			Y	Weekly	3 Weeks
13	Y	5/Day	Today	Y	Bi-Weekly	3 Days
14	N			Y	Daily	Yesterday
15	N			Y	Weekly	Week
16	N			Y	Weekly	2 Weeks
17	N			Y	4*Year	6 Months
18	N			N		
19	Y	10/Week	Today	Y	Weekly	2 Days
20	N			Y	Monthly	Month
21	Y	Weekly	2 Days	Y	Weekly	2 Days
22	N			Y	Weekly	3-4 Days
23	Y	Weekly	Today	Y	Weekly	Week
24	N			Y	Weekly	Yesterday
25	Y	Daily	Today	Y	Bi-Weekly	Month
26	N			Y	Monthly	3 Months
27	N			Y	Bi-Weekly	Week
28	N			N		
29	N			Y	Monthly	Week
30	Y	Daily	Today	Y	Daily	Week

31	N			Y	Weekly	Week
32	Y	Socially	2 Weeks	Y	Socially	Yesterday
33	N			N		
34	Y	Weekly	Week	Y	Weekends	Week
35	N			Y	Monthly	2 Weeks
36	N			Y	Monthly	Month
37	N/A	N/A	N/A	N/A	N/A	N/A
38	N			Y	Weekly	Week
39	N			Y	Daily	Yesterday
40	N			N		
41	N			Y	Weekly	Yesterday

Questionnaire Coffee & Medication:

PPt	Coffee	Amount	Last week	Meds	Amount	Last
1	Y	Daily	Yesterday	N		
2	Y	Daily	Today	N		
3	Y	Daily	Today	N		
4	Y	3/day	Yesterday	Y	Daily	Yesterday
5	Y	Weekly	Yesterday	N		
6	N			N		
7	Y	Weekly	2 Weeks	Y	Daily	Today
8	N/A	N/A	N/A	N/A	N/A	N/A
9	N/A	N/A	N/A	N/A	N/A	N/A
10	Y	Daily	Yesterday	N		
11	Y	Daily	Today	N		
12	N			N		
13	Y	2 Days	Today	Y	Weekly	Yesterday
14	N			N		
15	N			Y	Daily	Week
16	Y	Weekly	Yesterday	N		
17	N			N		
18	Y	Daily		N		
19	Y	Occasionally	Last week	N		
20	N			N		
21	N			N		
22	Y	Daily	Today	N		
23	Y	Weekly	Week	N		
24	Y	Daily	Yesterday	N		
25	Y	Weekly	Yesterday	N		
26	Y	Weekly	Week	N		
27	Y	Socially	Month	N		
28	N			N		
29	N			Y	Daily	Today
30	Y	Weekly	Today	N		

31	N			N		
32	N			N		
33	N			N		
34	y	Daily	Today	N		
35	N			N		
36	Y	Weekly	3 Days	N		
37	N/A	N/A	N/A	N/A	N/A	N/A
38	Y	Weekly	Week	N		
39	Y	Daily	Yesterday	N		
40	Y	Daily	Today	N		
41	Y	Daily	Yesterday	N		

EEG trial retention

Participant	Anger	Autonomy	Control	Joy	Neutral
1	27	16	18	21	22
2	20	21	19	22	17
3	13	15	17	12	10
4	3	6	11	13	13
5	22	20	17	19	22
6	22	20	23	21	27
7	4	13	18	17	11
8	Excluded prior to analysis				
9	Excluded prior to analysis				
10	16	21	22	17	21
11	27	29	26	29	30
12	0	3	1	1	1
13	11	13	12	19	15
14	17	19	25	20	15
15	27	24	27	28	30
16	22	15	18	21	17
17	28	24	24	26	22
18	19	21	21	21	18
19	22	18	19	21	19
20	17	15	19	21	24
21	17	23	20	26	24
22	28	29	27	27	29
23	26	24	27	32	24
24	30	32	29	31	30
25	30	31	30	29	32
26	23	28	31	22	29
27	16	15	13	17	16
28	25	21	19	24	24
29	30	25	26	23	28
30	23	26	24	24	20
31	31	31	32	29	31

32	28	28	31	31	30
33	29	31	29	28	27
34	26	27	26	21	25
35	11	19	14	12	13
36	30	24	27	31	29
37	Excluded prior to analysis				
38	25	28	28	26	28
39	27	29	18	23	27
40	27	26	25	22	25
41	24	27	25	26	26

Recorded demographics (excluding N/A's)

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	38	18	32	21.03	2.746
Valid N (listwise)	38				

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	21	55.3	55.3	55.3
	Female	17	44.7	44.7	100.0
	Total	38	100.0	100.0	

Retained after data cleaning

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Age	31	18	32	21.06	2.920
Valid N (listwise)	31				

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Male	18	58.1	58.1	58.1
	Female	13	41.9	41.9	100.0
	Total	31	100.0	100.0	

Stimuli validation (acoustics)

Anger

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	16	.218218088	.993708496	.58204004400	.241532698000
NrangeF0	16	.450171023	4.096932316	2.09767715700	1.446634117000
rangedB	16	19.96979005	54.00581067	36.8678740200	8.67250217500
EB_0_500	16	30.30592610	43.00985919	35.5081128400	3.83505392900
EB_0_1000	16	28.37827007	40.25331806	33.3630509700	3.53003935500

EB_500_1000	16	20.13599570	35.66826822	27.7202672000	3.54705238600
EB_1000_5000	16	8.868573132	28.155708070	18.03066476000	5.324438414000
EB__5000_8000	16	-4.270472893	18.643399230	4.95233854000	5.631224411000
duration	16	.980861678	1.771065760	1.42369472800	.253474966000
Speechrate	16	.162364755	.260090703	.20346427550	.026177843966
Valid N (listwise)	16				

Autonomy

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	32	.409922380	1.323804457	.77759677400	.251014153000
NrangeF0	32	.644106099	3.946395878	1.71445498900	.968601400000
rangedB	32	21.08605921	49.87022391	34.4358782800	7.63797074200
EB_0_500	32	26.62694068	42.12695219	35.1340898900	4.37373571700
EB_0_1000	32	25.86715854	39.85580175	32.9698119900	4.36790804300
EB_500_1000	32	15.46816742	37.99688025	26.7120158600	6.01651887000
EB_1000_5000	32	8.113732363	28.632895630	18.41315035000	5.484116473000
EB__5000_8000	32	-6.008930208	18.862613550	6.02175464000	6.523178169000
duration	32	.859954649	2.200136054	1.29361111100	.323456491000
Speechrate	16	.151924603	.274523810	.19367660475	.031935418731
Valid N (listwise)	32				

Control

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	16	.048365568	.781337488	.50238790500	.204661153000
NrangeF0	16	.318323792	4.013707633	1.63773140000	1.255918149000
rangedB	16	24.70580266	49.15957602	37.1122752500	6.52738477700
EB_0_500	16	31.73985124	40.54189373	36.0597656500	2.38760729400
EB_0_1000	16	30.55456435	37.86298911	34.2323858400	2.16359239600
EB_500_1000	16	27.19441833	34.08783997	30.4428015200	2.39156975400
EB_1000_5000	16	13.301217380	24.158630490	20.41838271000	3.018905996000
EB__5000_8000	16	-1.401825679	14.466517290	5.87751099400	5.307236603000
duration	16	.921405896	1.948299320	1.35353458100	.289484820000
Speechrate	32	.129866780	.291417234	.18701092884	.036781443084
Valid N (listwise)	16				

Joy

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	32	.542420083	1.981629219	1.26399065800	.338371729000
NrangeF0	32	.685599042	3.815029395	1.79605548700	.828286574000
rangedB	32	21.15214162	49.10261025	33.5116164500	5.11447573700
EB_0_500	32	32.51290212	48.27574793	41.4804391400	3.55072506200

EB_0_1000	32	30.81936187	45.43793030	39.5503298400	3.33187913900
EB_500_1000	32	25.77493991	42.57713073	34.1647770500	4.64644036700
EB_1000_5000	32	17.582274530	32.434141820	25.18126837000	3.998051352000
EB__5000_8000	32	-.923715999	20.850063840	12.12913551000	6.772776867000
duration	32	.834807256	1.562244898	1.05710600900	.155727204000
Speechrate	32	.129866780	.291417234	.18701092884	.036781443084
Valid N (listwise)	32				

Neutral

Descriptive Statistics[a]

	N	Minimum	Maximum	Mean	Std. Deviation
NmeanF0	32	.044384387	1.133750017	.40419584900	.296188196000
NrangeF0	32	.066873149	4.419642175	1.76379053600	1.481589423000
rangedB	32	18.21107355	42.75355579	34.1027083600	5.35153209900
EB_0_500	32	29.03567956	38.81653253	32.9224043100	2.40129479600
EB_0_1000	32	27.13676126	36.60398009	31.6206413500	2.56301638600
EB_500_1000	32	17.95947148	36.29441636	28.4801011600	4.68081253300
EB_1000_5000	32	11.994154770	26.940765740	18.24401835000	3.770275836000
EB__5000_8000	32	-8.483086416	19.132862270	-.52749556300	5.245822021000
duration	32	.769591837	1.602267574	1.11510700100	.242434329000
Speechrate	32	.122167153	.266224490	.18504400656	.033316079749
Valid N (listwise)	32				