Running head: ERP correlates of motivational communication

ERP correlates of Motivating Voices: Quality of Motivation and Time-Course Matters

Konstantina Zougkou[1], Netta Weinstein[2], Silke Paulmann[1]

[1]University of Essex, Department of Psychology and Centre for Brain Science

[2]Cardiff University, School of Psychology

Corresponding author:

Silke Paulmann

University of Essex

Department of Psychology & Centre for Brain Science

Wivenhoe Park

Colchester

CO4 3SQ

UK

paulmann@essex.ac.uk

Abstract

Here, we conducted the first study to explore how motivations expressed through speech are processed in real-time. Participants listened to sentences spoken in two types of well-studied motivational tones (autonomy-supportive and controlling), or a neutral tone of voice. To examine this, listeners were presented with sentences that either signaled motivations through prosody (tone of voice) and words simultaneously (e.g., "You absolutely have to do it my way" spoken in a controlling tone of voice), or lacked motivationally biasing words (e.g., "Why don't we meet again tomorrow" spoken in a motivational tone of voice). Event-related brain potentials (ERPs) in response to motivations conveyed through words and prosody showed that listeners rapidly distinguished between motivations and neutral forms of communication as shown in enhanced P2 amplitudes in response to motivational when compared to neutral speech. This early detection mechanism is argued to help determine the importance of incoming information. Once assessed, motivational language is continuously monitored and thoroughly evaluated. When compared to neutral speech, listening to controlling (but not autonomy-supportive) speech led to enhanced late potential ERP mean amplitudes, suggesting that listeners are particularly attuned to controlling messages. The importance of controlling motivation for listeners is mirrored in effects observed for motivations expressed through prosody only. Here, an early rapid appraisal, as reflected in enhanced P2 amplitudes, is only found for sentences spoken in controlling (but not autonomy-supportive) prosody. Once identified as sounding pressuring, the message seems to be preferentially processed, as shown by enhanced late potential amplitudes in response to controlling prosody. Taken together, results suggest that motivational and neutral language are differentially processed; further, the data suggest that listening to cues signaling pressure and control cannot be ignored and lead to preferential, and more in-depth processing mechanisms. **287 words**

Keywords: attitudes; prosody; tone of voice; language**;** motivation

Introduction

ERP correlates of Motivating Voices: Quality of Motivation and Time-Course Matters

Over the past decade, an increasing number of social and cognitive neuroscience studies have explored how social-affective intentions are perceived from speech and speech prosody in particular. Speech prosody is a term often used interchangeably with 'tone of voice', and can be described in terms of the supra-segmental features of speech: it relates to the fluctuations of various acoustic cues including pitch (high/low), loudness (increase/decrease), and temporal (fast/slow) features among others (e.g., Banse & Scherer, 1996). Within this line of research, much focus has been put on how listeners process emotional (see Paulmann, 2015 for review) and to a lesser extent attitudinal (see Mitchell & Ross, 2013 for review) vocal signals, as well as how words can convey emotional meaning (e.g., Kanske, Plitschka, Kotz, 2011; Schacht & Sommer, 2009; Schindler & Kissler, 2016). Yet, other interpersonally-laden experiences, such as motivation, have been heavily neglected in the endeavor to unravel how social communicative intentions are processed in the brain. Motivation reflects an *intrapersonal* experience (Deci & Ryan, 2000), which can be understood as the reason for action which energizes or directs behavior. Yet individuals also regularly attempt to motivate others, to elicit in them a drive to act – and this is an inherently *interpersonal* experience (Deci & Ryan, 1987; McClelland, 1987). The present study thus examines this interpersonal process, attempting to fill the gap in the literature by exploring how motivational intentions expressed through both lexical-semantic content and prosody (from now on referred to as "motivational speech") and prosody alone ("motivational prosody") are processed in real-time.

Control and Autonomy-Supportive Motivations

We focus on two qualities of motivation, or the impetus to action, which are

considered especially important in typical daily interactions where it is often the intention of one individual (the speaker) to shape others' behavior and energize them to action. For instance, consider the difference between saying to your partner "you really *have to* bring out the trash tonight" or saying, "if you would be willing, you *could* bring out the trash". The motivational qualities underlying these messages are quite distinct: in the first example, the partner is told what to do, while in the latter example, they are provided with the choice to act. Social psychology has a long tradition of examining how these two types of messages are experienced, employing self-determination theory (SDT; Deci & Ryan, 1985; Ryan & Deci, 2000), a theoretical framework for understanding human motivation which has been applied in experimental work examining strangers' interactions, in sports and education, parenting and close relationships, and clinical and health contexts, among others (e.g., Deci & Ryan, 2012; Deci, Ryan, & Guay, 2013; Hodgins, Brown, & Carver, 2007; Levesque & Pelletier, 2003; Radel, Sarrazin, & Pelletier, 2009; Silva, Marques, & Teixeira, 2014; Weinstein & Hodgins, 2009). SDT distinguishes two types of motivation: "controlling" (as in the first example above) and "autonomy-supportive" (as in the second example above) (Deci & Ryan, 1985; Ryan & Deci, 2000). Conveying controlling social messages is argued to drive others to action through coercion, or in order to conform with imposed expectations; in these cases the listener is left with a sense of pressure and lack of choice. In other words, the speaker is communicating how the listener should and must act. As messages are increasingly autonomy-supportive, listeners are provided with a clear sense of choice and volition, they can initiate an action on the basis of their own interests, beliefs, needs and because they value the outcome of the action. Thus, this type of motivation is experienced as self-endorsed.

Importantly, an individual subject to a controlling environment may respond in a very different way to the motivated activity than one who is autonomy-supported. A large body of work now indicates that individuals suffer long-term costs to well-being and health if they are

frequently exposed to controlling environments (see Deci & Ryan, 2012 for review; as well as Chen et al., 2015; Vansteenkiste & Ryan, 2013), and they show less interest and persistence in task-related behaviors over the long-term (reviews in Deci & Ryan, 2008; Ryan, 2012). In part, such costs are accrued because these individuals fail to process and differentiate the motivational messages, instead merely reacting to the immediate pressures without a sense of self-endorsement (Weinstein & Hodgins, 2009; Weinstein, Przybylski, & Ryan, 2013). On the other hand, autonomy-supported individuals show more awareness of new information in line with how self-relevant and desired this information is – in other words, they more deliberately and discriminately respond to new information (Brown & Ryan, 2003; Niemiec et al., 2010; Pennebaker & Chung, 2011; Weinstein et al., 2014), an important precursor for adaptive processing of information (Weinstein et al, 2013), and for selecting self-endorsed goals that reflect truly desirable ends (Milyavskaya, Nadolny, & Koestner, 2014). These findings all point to the importance of motivational qualities for how individuals respond to and process new information, and suggest that this process of responding has affective and behavioral implications for listeners.

Yet these findings and others employing a motivational approach have largely focused on the *perceptions* of individuals as being autonomy-supported or controlled, or in some cases work has identified which *words* individuals use to convey these motivational messages (e.g., Hodgins, Brown, & Carver, 2007; Levesque & Pelletier, 2003; Radel, Sarrazin, & Pelletier, 2009; Ryan, 2012; Weinstein & Hodgins, 2009). For example, words such as “should”, “must”, or “have to” are often used to communicate control, while words such as “choice”, “free”, and “your decision” have been used to activate autonomy-support in these paradigms. Until recently, research had failed to look *beyond* the use of words, but a new line of research focusing on prosody suggests that the two types of social motivations are expressed with particular prosodic patterns. This is the case even when the words used by

speakers do not bias towards one specific motivational reading (Weinstein, Zougkou, & Paulmann, 2014). In this work, controlling prosody is expressed with a lower pitch, louder tone, faster speech rate, and harsher voice quality when compared to autonomy-supportive prosody, for which pitch was found to be higher, speech rate slower, and voice quality milder (as reflected in a decrease in voice energy use). Building on this evidence that these two motivational climates are expressed with distinct prosody use, the current study will explore how these prosodic patterns are operationalized in real time to help shape our understanding of how social prosody and motivational prosody, in particular, is processed in the brain. Crucially, we will also investigate how processing of motivations as conveyed through prosody alone compares to processing motivations when conveyed through both lexical-semantic content and prosody.

A Multi-Stage Approach to Social Prosody

Here, we study perception of motivational speech by measuring event-related brain potentials (ERPs), which are sensitive to processes as they unfold over time. Collectively, ERP studies on *affective* speech have highlighted that acoustic signal processing is a multi-stage, rapid approach (see, e.g., Kotz & Paulmann, 2011; or Paulmann & Kotz, In Press, for reviews): initially, affective and non-affective signals need to be distinguished. This initial differentiation mechanism is believed to be triggered and linked to the extraction of meaningful acoustic cues (e.g., pitch, loudness, tempo). Next, the extracted information is combined to assess saliency and relevance before the social-affective meaning is fully determined. Currently underspecified, different contextual and individual factors are thought to modulate each of these different processing steps (e.g., Frühholz, Trost & Kotz, 2016; Kotz & Paulmann, 2011; Paulmann & Kotz, In Press; Schirmer & Kotz, 2006). Presumably, *attitudinal* (as opposed to *emotional*) signal processing follows similar steps. In fact, it has been speculated that initial stages, such as low level auditory processing and subsequent

binding of cues to form a prosodic composition, are comparable across different types of prosody, while later stages may engage a different underlying neural network (c.f. Mitchell & Ross, 2013), possibly because conveying attitudes at times requires more *subtl*e prosodic cue manipulations than conveying highly charged emotions.

Several neuro-biological markers, or ERP components, have been repeatedly linked to the different processing steps of social affective prosody processing. This supports the view that *multi-step* neural networks orchestrate vocal signal processing across both emotional and attitudinal stimuli (e.g., Frühholz et al., 2016; Kotz & Paulmann, 2011; Mitchell & Ross, 2013; Paulmann & Kotz, In Press; Schirmer & Kotz, 2006; Wildgruber et al., 2006). The earliest responding has been linked to the N1 component, a negative ERP deflection elicited around 100 ms after prosody onset, which is closely tied to the extraction of pitch, tempo, and loudness information of a signal. Enhanced N1 amplitudes have been reported for neutral when compared to angry prosody when conveyed through non-verbal vocalizations, suggesting that emotional information expressed through particularly salient acoustic cues is extracted rapidly by the listener (Liu et al., 2012).

The N1 is followed by the P2 component, which differs in amplitude between neutral and different types of emotionally (e.g., angry, sad, happy, surprised) intoned sentences (e.g., Paulmann & Kotz, 2008; Pell et al., 2015; Schirmer, Chen, Ching, Tan, & Hong, 2013). Very recent evidence suggests that mean amplitudes of the P2 component increases when listening to very confident versus not confident speakers (Jiang & Pell, 2015), or when listening to sarcastic versus non-sarcastic voices (Wickens & Perry, 2015). The P2 component thus seems to reflect very early tracking of social-affective saliency, including the speakers' psychological state. This is remarkable given that sentence prosody develops over time and, presumably, different meaningful acoustic cues become available throughout the sentence. Yet, listeners seem to be highly tuned towards evaluating socially relevant prosodic signals as

quickly as possible (e.g., Paulmann & Kotz, 2008; Schirmer et al., 2013). In fact, it can be argued that this very rapid appraisal of information is important given the likely impact that accurate perception of social prosody will have on listeners' behavior and social functioning. In the context of motivational speech, such rapid appraisal should be important for determining how and whether one should act or react, given such speech is designed to elicit specific responses from the listener.

Finally, early tracking of emotional and attitudinal signals is followed by further and deeper evaluations of social affective details as reflected in later, long lasting components such as the Late Positive Complex (LPC; Jiang & Pell, 2015; Paulmann, Bleichner & Kotz, 2013; Pell et al., 2015; Schirmer et al., 2013). Several previous studies describe differences in LPC amplitudes as a function of vocal emotions. For instance, angry sounding stimuli often elicit increased LPC amplitudes when compared to ERPs in response to sadness (e.g., Paulmann et al., 2013; Pell et al., 2015). More recent evidence also implies that this step is relevant in instances of processing speakers' attitudes such as when evaluating (in)sincerity. Listening to sincere compliments elicits higher late potential amplitudes as opposed to insincerely uttered compliments (Rigoulot, Fish, & Pell, 2014). Collectively, these studies suggest that integrative social-affective meaning evaluation processes are mirrored in increased LPC amplitudes. Thus, in line with multi-stage models of vocal signal processing (Frühholz et al., 2016; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), N1, P2, and LPC responses can be linked to three meaningful stages of attending to and comprehending social-affective components of speech.

Present Research

The present investigation aimed to understand how listeners process motivational prosody and speech, contributing to the growing body of literature on how social-affective intentions are perceived. Guided by motivational theory (Ryan & Deci, 2000), we explored

how two motivational qualities, autonomy-support and control, are processed in real-time. We followed a multi-stage neural network perspective which has previously been used to understand both emotions and attitudes (e.g., Schirmer & Kotz, 2006; Kotz & Paulmann, 2011; Mitchell & Ross, 2013; Frühholz et al., 2016), and expected that motivations would be distinguished from one another and from neutral speech at different time-points during processing.

While some existing attitudinal research has failed to find early differentiation in the P2 (Regel et al., 2011; Rigoulot et al., 2014; but see Wickens & Perry, 2015), and only report differently modulated late positive components, we expected neural responses to motivational prosody and speech would differ from responses elicited by neutral messages at both earlier and later time frames. Given their immediate relevance for action – for instance, it is important to realize quickly if you must immediately act in some way to satisfy the requirements of others around you - we hypothesized that motivational intentions communicated either through speech (prosody and word use), or prosody only, should be attended to in a rapid fashion, similar to what has been reported for vocal intentions signaling emotions (which would be reflected in enhanced N1 and P2 amplitude modulations for motivational as opposed to neutral signals). Yet, because they are intended to change the listeners' present and future behaviors, we expected they will also require enhanced processing at later stages (as reflected in enhanced LPC mean amplitudes in response to motivational prosody and speech).

Second, we expected that controlling communications, expressed through speech or only through prosody, would elicit enhanced P2 and LPC components when compared to ERPs in response to autonomy-supportive prosody and speech. This was hypothesized as controlling communications have been shown to differ from autonomy-supportive communications at the acoustic level, including speakers using a louder tone of voice for

controlling messages. The P2 component in particular has been linked to the extraction of salient acoustic cues (e.g., Paulmann & Kotz, 2008). Arguably, increase in loudness can be considered a salient cue for listeners. Moreover, the P2 component has been argued to be modulated by the relevance of an auditory stimulus (e.g., Kotz & Paulmann, 2011). Controlling prosody and speech are thought to push for immediate responses from listeners in specific ways (Deci, Spiegel, Ryan, Koestner, & Kauffman, 1982; Grolnick & Seal, 2008), providing such a case of relevance for listeners. In addition, some of our own preliminary findings also suggest that controlling motivational prosody and speech negatively affect listeners. For instance, we found that listening to controlling intonation predicts greater costs to subjective well-being, such as self-esteem, when compared to hearing an autonomy-supportive or neutral tone of voice (Weinstein, 2016). Similarly, we expect enhanced LPC amplitudes in response to controlling prosody and speech given the presumed link between LPC responses and enhanced social-affective meaning processing (e.g., Pell et al., 2015).

Methods

Participants

Twenty-two native English speakers were recruited from the University of Essex to take part in the study. Of these, two were excluded because one was Australian and the other withdrew halfway through the session, leaving a total of 20 British participants. None of the participants reported taking medication for psychopathology or mood disorders. In addition, none of the participants reported any hearing difficulties, and all had right hand dominance as assessed by an adapted version of Edinburgh Handedness Inventory (Oldfield, 1971). Of the final sample, 8 were male and 12 female, with a mean age of M = 19.7 years (range: 18-26 years).

Materials

Speech stimuli: Sentences were taken from a previously validated pool of materials (Weinstein et al., 2014). Two male (both 21 years old) and two female speakers (19 and 27 years old) were selected based on these validation ratings. Participants were presented with five types of sentences, reflecting prosody and speech conditions: 1) sentences expressing autonomy-support through prosody and word use (e.g., *You may do this if you choose* spoken in an autonomy-supportive prosody; called autonomy-supportive speech); 2) sentences expressing control through prosody and word use (e.g., *You have to do this now* spoken in a controlling prosody; called controlling speech; 3) semantically neutral sentences intoned with a non-motivational, neutral tone of voice (e.g., *You are quite tall for your age*; called neutral speech); 4) semantically neutral sentences spoken in an autonomy-supportive prosody (referred to as autonomy-supportive prosody); and 5) semantically neutral sentences spoken in a controlling prosody (referred to as controlling prosody). Thus, in these sentences motivation could be expressed through lexical-semantic and prosodic information simultaneously, or through prosody alone. To manipulate prosody and speech, one-hundred different sentences were selected from the database for each condition, with each speaker occurring equally often (i.e., each speaker intoned 25 sentences for each condition). We chose to present one-hundred stimuli per condition to ensure that a comparable number of data points as used in previous ERP prosody studies (e.g., Pell et al., 2015; Rigolout et al., 2014; Paulmann et al., 2013) would enter statistical analysis.

The number of words per sentence was matched between conditions, and both sets of sentences ranged in length from 3 to 9 words. The mean duration (sec) was similar across conditions (autonomy-supportive speech: 1.63, SD = 0.25; autonomy-supportive prosody: 1.33, SD = 0.22; controlling speech: 1.87, SD = 0.37; controlling prosody: 1.62, SD = 0.36; neutral speech: 1.59, SD = 0.37). Stimuli were significantly different between conditions in terms of pitch: $F(4, 495) = 11.18$, $p = .001$, amplitude: $F(4, 495) = 73.21$, $p = .001$, and

speech rate: $F(4, 495) = 20.90$, $p = .001$. Table 1 shows means and standard deviations for the three acoustic indicators across all conditions. Follow-up analysis showed both autonomy supportive speech and prosody differed from controlling speech and prosody in terms of pitch (*ps* < .05, $p = .06$ between autonomy-supportive speech and controlling prosody), amplitude (*ps* < .001) and speech rate (*ps* < .001). Autonomy-supportive speech and prosody were acoustically different from neutral speech in terms of pitch (*ps* < .001) and speech rate (*ps* < .001), but not amplitude ($p = .99$). Controlling speech and prosody were not different from neutral speech in terms of pitch ($p = .09$ and $p = .07$, respectively) or speech rate ($p = .92$ and $p = .90$, respectively), but in terms of amplitude (*ps* < .001). Autonomy-supportive speech did not differ from prosody on any of the three acoustic parameters (pitch $p = .97$, amplitude $p = .98$ and speech rate $p = .88$). Similarly, controlling prosody sentences were not acoustically different from controlling speech sentences in terms of pitch ($p = 1.00$), amplitude ($p = .94$) or speech rate ($p = .43$). See Figure 1 for example spectrograms.

In an initial validation of these stimuli, autonomy-supportive speech and autonomy-supportive prosody were seen as more supportive of choice ($M = 4.02$, $SD = 0.40$ and $M = 3.45$, $SD = 0.39$, respectively) as compared to pressuring ($M = 1.82$, $SD = 0.39$ and $M = 1.82$, $SD = 0.33$, respectively) in a previous five-point scale validation of our sentences (Weinstein et al., 2014). Controlling speech and controlling prosody sentences were rated as more pressuring ($M = 4.06$, $SD = 0.38$ and $M = 3.45$, $SD = 0.62$, respectively) and not supportive of choice ($M = 1.69$, $SD = 0.29$ and $M = 2.31$, $SD = 0.51$). Neutral speech sentences were not seen to be particularly choice-promoting ($M = 2.90$, $SD = 0.35$) or pressuring ($M = 2.20$, $SD = 0.44$).

Table 1

Results from Acoustical Analyses Comparing All Tested Conditions. Analyses Were Conducted with Praat (Boersma & Weenink, 2013).

	F0 mean (*SD*)	dB mean (*SD*)	speech rate mean (*SD*)
Autonomy-supportive speech	201.44 (49.00)	52.94 (3.04)	0.20 (0.03)
Autonomy-supportive prosody	205.99 (59.64)	53.23 (3.04)	0.19 (0.03)
Controlling speech	182.32 (43.97)	59.04 (4.17)	0.23 (0.04)
Controlling prosody	183.05 (47.51)	58.63 (4.06)	0.24 (0.06)
Neutral	164.72 (46.51)	53.09 (3.94)	0.23 (0.05)

Note. F0 = fundamental frequency; dB = mean intensity as measured in decibel; speech rate = seconds per syllable.

Procedure

EEG recordings were acquired in a sound attenuated booth. All participants were seated approximately 100 cm away from a computer screen. Materials were presented using SuperLab 5 in a fully randomized order. Participants were asked to listen to materials carefully as they would be filling in a short questionnaire related to the sentences at the end of the session. This task was used to ensure that participants paid attention to the materials without explicitly focusing on the motivational qualities of the sentences, and as such responses to this task were not analyzed. Materials were distributed over five blocks with 100 trials each. An experimental trial operated as follows: participants first fixated on a cross presented in the middle of a computer screen. Three-hundred milliseconds later, a vocal stimulus was presented via speakers located to the left and right side of the monitor while the fixation cross remained on screen. An inter-stimulus interval (ISI) of 1500ms preceded the next stimulus presentation. Before the start of the experiment, three practice trials were presented which familiarized participants with the procedure.

ERP recording

The EEG was recorded from 63 Ag–AgCl electrodes mounted on a custom-made cap (waveguard) according to the modified extended 10–20 system using a 72 channel Refa amplifier (ANT). Signals were recorded continuously with a band pass between DC and 102

Hz and digitized at a sampling rate of 512 Hz. Electrode resistance was kept below 7 KΩ. The reference electrode was placed on the left mastoid and data was re-referenced offline to averaged mastoids. Bipolar horizontal (positioned to the left and right side of participants' eyes) and vertical EOGs (placed below and above the right eye) were recorded for artifact rejection purposes using disposable Ambu Blue Sensor N ECG electrodes. CZ served as a ground electrode.

Data Analysis

Data were filtered off-line with a band pass filter set between 0.01 and 30 Hz and a baseline correction was applied. For each ERP channel, the mean of our baseline time-window (-200 to 0 ms) was subtracted from the averaged signal. All trials containing muscle or EOG artifacts above 30.00 μV were automatically rejected using EEProbe Software. Additionally, EEG data were also visually inspected to exclude trials containing additional artifacts and drifts. In total, 21% of data were rejected (range for different conditions: 20% – 22%).

After data cleaning, separate ERPs for each condition at each electrode-site were averaged for each participant with a 200ms pre-stimulus baseline for epochs lasting 1000ms post-stimulus onset. Epochs were time-locked to sentence onset of stimuli. Selection of time windows for ERP mean amplitudes was guided by previous research in the emotional prosody literature (Pell et al., 2015; Paulmann et al., 2013). Three components were of a priori interest: N1, P2, and late component. The N1 component of the averaged data showed a mean peak latency of 130 ms; the time window of interest was set to 80-170 ms (similar to Pell et al., 2015). The P2 time window was set between 170 and 230ms (peak latency for the averaged data: 200 ms; Paulmann et al, 2013) and to explore late component differences, we set the time window between 350 and 600ms to cover a relatively wide temporal breadth of this later effect (see Olofsson, Nordin, Sequeira, and Polich, 2008 for summary of previously

explored long latency time windows). Electrode-sites were grouped according to hemisphere (left/right) and region (frontal, central, parietal), with midline electrodes being analyzed separately: left frontal electrode-sites (F5, F3, FC5, FC3); right frontal electrode-sites (F6, F4, FC6, FC4); left central (C5, C3, CP5, CP3); right central (C6, C4, CP6, CP4); left posterior (P5, P3, PO7, PO3); right posterior (P6, P4, PO8, PO4); and midline sites (Fz, Cz, CPz, Pz). This electrode grouping approach allowed us to cover a broad scalp range to explore potential topographical differences.

Three analyses were conducted for each time-window using the "proc glm" function in SAS 9.4. These analyses looked at prosody effects only (including the conditions of autonomy-supportive prosody, controlling prosody, and neutral prosody), speech effects only (including the conditions of autonomy-supportive speech, controlling speech, and neutral speech), and prosody versus speech effects (including the conditions of autonomy-supportive prosody vs. autonomy-supportive speech, and controlling prosody vs. controlling speech). Mean amplitudes for each time-window were analyzed separately with repeated-measures analyses of variance (ANOVAs) treating *condition*, *hemisphere*, and *region* as within-subjects factors. Main effects and interactions involving the critical *condition* factor at $p < .05$ were followed up by simple effects analyses and pairwise comparisons (we also report effects approaching significance ($p \leq .08$) to inform readers about emerging patterns). The Greenhouse-Geisser correction was applied to all repeated measures with greater than one degree of freedom in the numerator.

Results

ERP Data

Figures 2 and 3 display ERP waveforms in response to motivational prosody (Figure 2A) and motivational speech (Figure 2B). Figures 2C and 2D present a comparison of motivational prosody and motivational speech effects. Figure 3 displays topographical maps

for the effects. For all conditions, an early N1 modulation was followed by a broadly distributed P2 component, followed by a longer lasting potential. Differences between conditions appeared most visible at central electrodes-sites.

N1 (80-170 ms)

Prosody. To explore brain responses in response to motivational prosody, we compared ERPs in response to autonomy-supportive prosody, controlling prosody, and neutral speech in three different time windows. We report results from analyses including hemisphere (left/right) and region (frontal, central, posterior) first, followed by the analyses using midline electrodes.

In the early time-window, no main effects or interactions involving the critical *condition* factor reached or approached significance (all $Fs < 2.3$, $ps > .12$). Similarly, at midline electrodes, no *condition* effect was found ($F = 1.21$, $p = .30$). Thus, there is no evidence for differences in ERP amplitudes in response to motivational prosody at this very early time-window.

Speech. To investigate processing of motivations communicated through both lexical-semantic and prosodic cues, we compared ERPs in response to autonomy-supportive, controlling, and neutral speech. No effects involving the factor *condition* were found in either of our analyses (all $Fs < 2.2$, $ps > .14$).

Prosody versus Speech. ERPs in response to autonomy-supportive prosody and autonomy-supportive speech, as well as controlling prosody and controlling speech, were compared to investigate differences in brain responses of motivational speech and prosody. This analysis revealed a *condition* X *region* interaction, $F(6, 108) = 4.02$, $p = .01$. Analyses by *region* revealed an effect of *condition* for the posterior region only, $F(3, 54) = 2.98$, $p = .047$. N1 amplitudes in response to autonomy-supportive prosody were less negative than those in response to autonomy-supportive speech, $F(1, 18) = 5.14$, $p = .04$, $d = .36$. No other

effects were in evidence (See Table 2 for this and future prosody X speech interaction effects). No effects were found at midline electrode-sites ($F = 1.4$, $p > 0.27$).

Table 2

Significant Contrasts For Direct Comparisons Between Speech and Prosody Conditions

Time Window	Contrast	Region	Prosody vs. Speech
N1	A-S vs. A-S	posterior	5.14 .04
	A-S vs. A-S	BH	5.42 .032
	A-S vs. A-S	RH	7.47 .014
	A-S vs. A-S	LH	.
	A-S vs. A-S	ML	4.44 .05
P2	C vs. C	BH	3.57 .075
	C vs. C	RH	.
	C vs. C	LH	3.53 .077
	C vs. C	ML	5.21 .035
	A-S vs. A-S	BH	4.30 .053
late potential	A-S vs. A-S	ML	3.42 .081
	C vs. C	BH	.
	C vs. C	ML	.

Notes. *F* and *p* values presented in top and bottom cells, respectively. A-S = autonomy-support; C = control; N = neutral; BH = both hemispheres; ML = midline; RH = right hemisphere; LH = left hemisphere.

P2 (170-230 ms)

Prosody. Figure 2 displays ERP effects at selected electrode-sites and Figure 3 shows topographical distribution. Focusing on electrode-sites in either hemisphere or region of interest, a main effect of *condition*, $F(2, 36) = 4.21$, $p = .02$, suggested that listening to different motivational tones elicited differently modulated ERP amplitudes. Planned follow-up comparisons showed that controlling prosody elicited increased P2 amplitudes compared to autonomy-supportive prosody, $F(1, 18) = 4.38$, $p = .05$, $d = .25$ (Table 3). Similarly, P2 amplitudes in response to controlling prosody were more positive than those in response to neutral speech, $F(1, 18) = 8.21$, $p = .01$, $d = .42$. Responses to autonomy-supportive prosody and neutral speech did not differ significantly for the electrode-sites included in this analysis ($p > .50$).

Looking at midline electrodes only, a main effect of *condition* was again found, $F(2, 36) = 6.60$, $p = .004$. Follow-up comparisons showed enhanced P2 amplitudes in response to controlling as compared to autonomy-supportive prosody, $F(1, 18) = 5.71$, $p = .03$, $d = .26$, and as compared to neutral speech, $F(1, 18) = 12.31$, $p = .003$, $d = .35$.

Speech. Looking at left and right-hemisphere electrode-sites, there was a main effect of *condition*, $F(2, 36) = 6.47$, $p = .004$, revealing differences in the P2 amplitudes between listeners' responses to motivational speech. Follow-up comparisons showed that autonomy-supportive speech elicited more positive P2 amplitudes than neutral speech, $F(1, 18) = 4.94$, $p = .04$, $d = .39$. Similarly, controlling speech led to more enhanced P2 amplitudes than neutral speech, $F(1, 18) = 13.59$, $p = .002$, $d = .48$.

The main effect was qualified by a *condition* X *hemisphere* interaction, $F(2, 36) = 3.83$, $p = .03$. Follow-up comparisons by hemisphere showed small effect sizes for the contrasts between autonomy-supportive speech and controlling speech at left hemisphere electrodes, $F(1, 18) = 3.34$, $p = .08$, $d = .31$, as well as between autonomy-supportive and neutral speech, $F(1, 18) = 3.37$, $p = .08$, $d = .32$. Further, results revealed a significant

difference between controlling and neutral speech, an effect with a moderate effect size, $F(1, 18) = 13.89$, $p = .002$, $d = .51$. P2 amplitudes in response to controlling speech were most positive, followed by autonomy-supportive speech, and neutral speech. Contrasts at right-hemisphere sites revealed P2 differences between autonomy-supportive and neutral speech, $F(1, 18) = 6.13$, $p = .02$, $d = .43$, as well as between controlling and neutral speech, $F(1, 18) = 11.67$, $p = .003$, $d = .43$. Again, ERPs in response to controlling speech were most positive, followed by autonomy-supportive and neutral speech.

Analyses for midline electrodes mirrored these results. A *condition* effect was found, $F(1, 18) = 10.99$, $p < .001$, and follow-up contrasts showed a rather small effect size for the contrast between ERPs in response to autonomy-supportive and controlling speech, $F(1, 18) = 3.66$, $p = .07$, $d = .25$. However, significant differences between autonomy-supportive and neutral speech were found, $F(1, 18) = 7.02$, $p = .02$, $d = .38$. Differences were also found between controlling and neutral speech, $F(1, 18) = 23.79$, $p < .001$, $d = .53$. Listening to controlling speech led to the most positive P2 amplitudes, followed by autonomy-supportive and neutral speech.

Table 3

Significant Contrasts for Both Speech and Prosody Conditions

Time Window	Contrast	Region	Prosody	Speech
P2	A-S vs. N	BH	.	4.94 .04
	A-S vs. N	ML	.	7.02 .016
	A-S vs. N	LH	.	3.37 .08
	A-S vs. N	RH	.	6.13 .024
	C vs. N	BH	8.21 .0103	13.59 .002

	C vs. N	ML	12.31 .0025	23.79 .0001
	C vs. N	LH	. 	13.89 .002
	C vs. N	RH	. 	11.67 .003
	A-S vs. C	BH	4.38 .0508	.
	A-S vs. C	ML	5.71 .028	3.66 .072
	A-S vs. C	LH	. 	3.34 .084
	A-S vs. C	RH	. 	.
	A-S vs. N	BH	. 	.
	A-S vs. N	ML	. 	.
late potential	C vs. N	BH	3.33 .0848	9.09 .007
	C vs. N	ML	3.48 .0784	13.47 .002
	A-S vs. C	BH	4.67 .0445	.
	A-S vs. C	ML	6.81 .0177	4.99 .038

Notes. *F* and *p* values presented in top and bottom cells, respectively. A-S = autonomy-support; C = control; N = neutral; BH = both hemispheres; ML = midline; RH = right hemisphere; LH = left hemisphere.

Prosody versus Speech. A main effect of *condition*, $F(3, 54) = 4.13$, $p = .01$, indicated enhanced P2 amplitudes in response to autonomy-supportive speech when compared with autonomy-supportive prosody, $F(1, 18) = 5.42$, $p = .03$, $d = .35$. The contrast between ERPs in response to controlling speech and controlling prosody was not significant, $F(1, 18) = 3.57$, $p = .08$, $d = .20$. The non-significant interaction between *condition* and *hemisphere*, $F(3, 54) = 2.65$, $p = .07$, was followed-up by hemisphere and revealed a small effect size for the contrast between controlling speech and controlling prosody at left hemisphere electrode sites, $F(1, 18) = 3.53$, $p = .08$, $d = .23$, suggesting more positive P2 amplitudes for controlling speech. At right hemisphere sites, the contrast between autonomy-

supportive prosody and autonomy-supportive speech reached significance, $F(1, 18) = 7.47$, $p = .01$, $d = .39$, showing more positive P2 effects for autonomy-supportive speech. The contrast between controlling prosody and controlling speech did not do so at right hemisphere sites ($F = 2.87$, $p = .10$).

At midline electrode sites, there was a main effect of *condition*, $F(1, 18) = 5.71$, $p = .003$. Pairwise comparisons showed a difference between autonomy-supportive speech and autonomy-supportive prosody, $F(1, 18) = 4.44$, $p = .049$, $d = .27$, as well as between controlling prosody and speech, $F(1, 18) = 5.21$, $p = .03$, $d = .18$. In both instances, ERPs in response to speech were more positive than those in response to prosody.

Late Potential (350-600 ms)

Prosody. Analyses for the late potential effect showed no *condition* omnibus effect, $F(2, 36) = 2.73$, $p = .08$. Exploratory follow-up comparisons informed by our hypotheses stated above showed more positive-going amplitudes in response to controlling prosody when compared to autonomy-supportive prosody, $F(1, 18) = 4.67$, $p = .045$, $d = .37$, but only a small effect size was found when comparing controlling prosody with neutral prosody, $F(1, 18) = 3.33$, $p = .08$, $d = .33$. No differences were found between autonomy-supportive and neutral prosody at these electrode-sites in this later time-window.

Focusing on midline electrode-sites only, a main effect of *condition*, $F(2, 36) = 3.59$, $p = .045$, indicated that controlling prosody elicited more positive-going late potentials than autonomy-supportive prosody, $F(1, 18) = 6.81$, $p = .02$, $d = .41$. The contrast between controlling prosody and neutral speech showed no significant difference, $F(1, 18) = 3.48$, $p = .08$, $d = .33$. Neutral speech responses did not differ from ERPs linked to autonomy-supportive prosody.

Speech. Analyses revealed a main effect of *condition*, $F(2, 36) = 4.12$, $p = .03$. Follow-up contrasts showed more positive-going ERP amplitudes for controlling as

compared to neutral speech, $F(1, 18) = 9.09$, $p = .007$, $d = .52$. No other effects were significant, $ps > .10$.

Analyses using midline electrodes also identified a main effect of *condition*, $F(2, 36) = 6.53$, $p = .004$. Planned post-hoc comparisons showed differences between autonomy-supportive and controlling speech, $F(1, 18) = 4.99$, $p = .04$, $d = .45$, as well as between controlling and neutral speech, $F(1, 18) = 13.47$, $p = .002$, $d = .55$. ERP waveforms for controlling speech were most positive, followed by autonomy-supportive and neutral speech.

Prosody versus Speech. A main effect of *condition* was observed when looking at ERPs in the six regions of interest, $F(3, 54) = 3.23$, $p = .045$. Pairwise comparisons between autonomy-supportive speech and prosody showed a difference, $F(1, 18) = 4.30$, $p = .05$, $d = .35$, such that more positive-going amplitudes were observed for speech as opposed to prosody. A similar effect was not found when comparing amplitudes in response to controlling speech and prosody ($F = 1.6$, $p = .22$).

At midline electrodes, a significant *condition* effect was found, $F(3, 54) = 5.87$, $p = .003$. Pairwise comparisons revealed a small effect size for the contrast between autonomy-supportive prosody and speech, $F(1, 18) = 3.42$, $p = .08$, $d = .27$. Amplitudes for prosodic stimuli were most positive. No other effects were significant.

Discussion

This is the first study to investigate different ERP markers of vocal motivational signal processing. In particular, we explored how two motivational qualities, namely autonomy-support and control, are processed in real-time when communicated through prosody only, as well as when communicated through a combination of sentence content and prosody. Building on work studying affective (Liu et al., 2012; Paulmann & Kotz, 2008; Schirmer et al., 2013; Wickens & Perry, 2015) and attitudinal speech (Jiang & Pell, 2015; Rigoulot et al., 2014; Wickens & Perry, 2015), we focused on three different processing

stages: processing of sensory information (N1), differentiation of salient and non-salient cues (P2), and more fine-grained analyses of stimuli (late component). Examining these different stages allowed us to describe the time-course underlying vocal motivational signal processing. Our findings indicated that vocal motivations are processed rapidly, similar to emotional and some attitudinal aspects of spoken language (e.g., Jiang & Pell, 2015; Paulmann & Kotz, 2008; Paulmann et al., 2013; Schirmer et al., 2013). Specifically, we observed that vocal motivations are processed within 200 ms of speech onset, but the two motivational climates of interest were also processed differently at later points in time, suggesting that listeners take into account the motivational intention expressed by the speaker at various stages during on-line speech processing. Taken together, results showed that motivational qualities as conveyed through the voice are processed differently at distinct time-points.

Early Processing of Motivational Prosody. In an attempt to outline the time-course underlying motivational prosody, we explored the N1 component, which was of interest given some evidence in the literature that emotions expressed vocally can lead to differences in N1 amplitudes (Pell et al., 2015; Liu et al., 2012). Contrary to findings from studying emotions, the current data provide no evidence for a similar very early differentiation between the two different motivational qualities; that is, no processing differences were found between controlling and autonomy-supportive stimuli. The absence of an effect at this early stage is interesting because there is some evidence that N1 amplitudes are dependent on the saliency of information provided (Liu et al., 2012). If true that N1 amplitudes are modulated by saliency of cues, it can be speculated that motivational prosody either lacks these "saliency" cues (e.g., certain pitch, tempo, and intensity combinations), or at least that the cues and/or specific cue configurations are modulated in a less pronounced way when expressing motivations. This conclusion is in line with the hypothesis that conveying

attitudes relies on "intentionally controlled processes" (Mitchell & Ross, 2013), in which the speaker, in this case, the motivator, actively modulates voice to elicit a certain response. These intentionally modulated communications might diverge from the much larger body of work on emotional expressions both in terms of the functional neuro-anatomy (Mitchell & Ross, 2013) and timing (Wickens & Perry, 2015), and they may result in less "prototypical" expressions. As such, they may lack some of the more prominent, or salient, acoustic features of typical emotional vocal expressions (such as using high pitch and intensity to signal anger) and rely on more subtle, more varied, prosodic cue manipulations. This idea is supported by earlier work showing that motivational speakers may vary in how they use pitch to convey autonomy-support and control, while emotions employ pitch in a robust and consistent way (Weinstein et al., 2014).

Conceptually and operationally, we also may expect emotions and motivations to differ. For example, an angry person can still use autonomy-supportive language, and a person can be controlling but calm. Emotional communications are in most cases intended to express a feeling, not to inspire others to action as in the case of motivational communications. Despite this, some theorists have treated the two constructs as interchangeable (e.g., Weiner, 1985), and in these cases the two literatures are not clearly defined or discriminated. The current data suggest then that control and anger do not engage in the same processing mechanisms, given early N1 differences in response to angry prosody versus a neutral comparison (Liu et al., 2012; also note that visual inspection of data by Paulmann, Ott, & Kotz, 2011, show the same N1 pattern) which is not found for controlling prosody. The discrepancy between the present findings and those which focus on anger thus suggest that distinct social-affective prosody and speech patterns are evaluated differently within 100 ms of speech onset, a finding that multi-stage models of social-affective signal processing will have to address in the future. These findings are important because they

suggest that motivational prosody processes are distinct from the much larger body of work on emotions. Given motivational speech seems to be processed differently in the brain, it is important to continue to study their independent contributions to listeners' experiences and behaviors.

Discriminations in Both Motivational Prosody and Speech Messages. The first differentiation between controlling and autonomy-supportive vocal communications was visible in the P2 component. Our results showed more enhanced P2 amplitudes in response to controlling as compared to autonomy-supportive and neutral prosody. The finding that controlling prosody is particularly attended to goes well with accounts that report preferential processing of salient attitude-revealing auditory stimuli (e.g., when a speaker's high arousal or confidence is conveyed; Jiang & Pell, 2015; Paulmann et al., 2013). In fact, we have previously argued that the P2 component is closely linked to the initial evaluation that incoming speech prosody is significant (e.g., Paulmann & Kotz, 2008; Paulmann et al., 2013), as potentially immediately concerning stimuli (e.g., those expressing anger or disgust or which could otherwise be considered action-relevant) elicit enhanced P2 amplitudes (Jessen & Kotz, 2011; Paulmann et al., 2013). Our findings therefore suggest that a controlling tone of voice calls for more immediate attention or reaction. This finding is also in line with the motivation literature, which suggests that controlling styles are used to affect a stronger or more instant reaction from others (Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Gagné & Deci, 2005), for example, in the case of managers who hope their workers will meet ambitious, immediate deadlines, or parents whose toddlers are about to touch a hot stove.

Unlike controlling prosody, autonomy-supportive prosody did not trigger enhanced attentional processes unless prosody was also paired with autonomy-supportive language. Results from pairing lexical-semantic and prosodic cues showed that listeners not only

distinguished early between controlling and neutral sentences, but also differentiated between autonomy-support and neutral within 200 ms of sentence onset. This suggests that as soon as autonomy-supportive messages (here motivational content and prosody) are conveyed with enough salient cues, they are processed differently than non-motivating messages (neutral content and prosody). The differential findings for autonomy-supportive speech, as opposed to autonomy-supportive prosody, are interesting in light of theory suggesting that autonomy-support is a way to motivate individuals by allowing them to make personally meaningful choices and pursue self-endorsed ends. However, applying such a strategy requires a more nuanced and meaningful invitation to self-exploration and self-direction (e.g., Ryan & Deci, 2000). Here, listeners seemed only to respond to autonomy-supportive motivating speakers when those speakers explicitly communicated the sense of choice through their use of words (such as, 'if you choose') *and* through their tone of voice. It may then be that listeners benefited from receiving a clear invitation to self-directed action, but did not react when hearing a supportive tone of voice which was absent of such a motivational meaning. In terms of the motivational literature, the more explicit invitation might be expected to result in a higher sense of well-being (e.g., Deci & Ryan, 2008; Ryan & Deci, 2000), more exploration and curiosity (Niemiec & Ryan, 2009; Roth, Assor, Niemiec, Ryan, & Deci, 2009), and more positive relational and performance outcomes (Weinstein, Hodgins, & Ryan, 2010). Hence, enhanced P2 components might reflect a more adaptive attention to motivating speakers that is distinct from the more immediate and compulsive reactions to controlling tones of voice, but this expectation would need to be tested in future research.

Either way, both in the cases of controlling prosody and speech, and in the cases of autonomy-supportive speech, the P2 component in response to motivational speech was more enhanced than the neural response to non-motivating speech. This finding fits well with previous observations in the literature which argued that the P2 is linked to relevance

(Paulmann et al., 2013). Here, we can extend this view as the data suggest that hearing input that calls for action or participation through prosody and content (e.g., "you can" [do this] vs. "you ought" [to do this]) engages similar processing mechanisms, but differs from receiving input in which no such action is required according to the tone of voice and words used by the speaker (e.g., "why don't" [you do this]). These data then fit well with previous reports showing differences in listeners' ERP responses to materials that convey emotional information through prosody only, or through prosody and lexical-semantic information (e.g., Kotz & Paulmann, 2007; Paulmann & Kotz, 2008b; Paulmann Jessen, Kotz, 2012). In these prior studies, listeners' expectancies generated by relatively short (e.g., "He has"/"She has" or the equivalent pseudo-language versions such as "Hung set") auditory input were violated in two ways: either, listeners were presented with a sentence ending that did not match the neutral prosody *and* neutral semantics (combined prosodic/semantic condition), or with sentence endings that only violated the prosodic expectancy, but not the semantic expectancy. While combined expectancy violations were detected earlier and indexed through a negative ERP component, detection of prosody-only violations were indexed through the prosodic-expectancy-positivity (PEP). Thus, listeners quickly built up expectancies about how a sentence would continue (both with regard to prosody and content information). Similarly here, the information provided through motivational content and prosody in our speech conditions were processed rapidly in a combined fashion and ERP effects were distinguishable from the prosody-only condition. In short, the P2 results observed for the data here nicely mirror and expand previous findings from the emotional and attitudinal prosody and speech literature (e.g., Jiang & Pell, 2015; Paulmann & Kotz, 2008; Paulmann et al., 2013; Schirmer et al., 2013). They show that a speakers' social intention (e.g., to convey confidence, to motivate) is assessed rapidly during on-line speech processing (e.g., Jiang & Pell, 2015). They also lend further support to previous claims that lexical-semantic cues can

pre-dominate prosody (e.g., Besson, Magne, Schön, 2002; Kotz & Paulmann, 2007; Paulmann, Jessen, Kotz, 2012; Paulmann, Titone, Pell, 2012). Although prosody matters, as soon as lexical-semantic cues are available as well, the combination of semantics and prosody seems to matter more. The finding that messages, at an early stage of processing, are responded to differently depending on whether the social-affective or motivational intention is conveyed through prosody only, or reinforced through the words used, is important for neuro-cognitive models of social signal processing. Currently, it is assumed that prosodic and semantic information is integrated within 400 ms of speech onset; however, the data here suggest that combined processing of cues can occur earlier (see also Paulmann, Jessen, & Kotz, 2012 for similar findings using a different experimental paradigm).

It should be noted that for the speech materials used here, different sentence onsets were used which, presumably, aided participants in predicting upcoming lexical-semantic information (see e.g., Laszlo and Federmeier (2009) for review of studies exploring prediction of upcoming words). While ecologically valid, this procedure might have introduced more variability in ERPs in response to motivational speech than what would have been found if all speech materials had started with the same words (e.g., you will [have to do it my way] vs. you will [be given an option] vs. you will [experience an event]). However, given that previous research on emotional prosody has reported comparable P2 effects for materials where the exact same sentence onset words were used (e.g., Kotz & Paulmann, 2007; Paulmann & Kotz, 2008), or those where different sentences were presented across different conditions (e.g., Pell et al., 2015; Paulmann & Uskul, 2017), it seems unlikely that differences in sentence onsets across the speech conditions are the driving force underlying the observed ERP effects.

Continued Attention to Non-Relevant Sentences: It's All About Control!

In line with multi-stage models of social-affective signal processing (e.g., Frühholz, Trost, Kotz, 2016; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006), our data also suggest that the rapid encoding of motivationally relevant information is followed by a later, more cognitively driven elaboration of motivational characteristics. Specifically, an enhanced late potential was observed in response to controlling as opposed to autonomy-supportive prosody and speech.

While we did not expect the finding that control, but not autonomy-support, would show late potential differences when compared to neutral speech, these findings in light of the methodology used here may be in line with theory and past research which describes processing of information in controlling and autonomy-supportive conditions. We argued above that the P2 primarily reflects processes that link to evaluating whether or not a stimulus is of relevance to the listener (e.g., to act/not to act). However, late potentials, that is effects occurring later than 300 ms after stimulus onset, have been argued to reflect continuous analysis of stimuli, particularly focusing on the continued monitoring of motivationally or emotionally relevant features (c.f. Paulmann & Kotz, In Press, for auditory emotion processing and c.f. Olofsson et al., 2008 for a review of visual emotion studies). Specifically, different, long latency components displaying different polarities (i.e. positive and negative ERPs such as N300, N400, late negative component, LPP) have been associated with enhanced and more sustained encoding of emotional and motivational attributes (c.f. Kotz & Paulmann, 2011 for review). In other words, for the data presented here, this would suggest that the analysis of acoustic cues which - if appropriately combined - signal a sense of the speaker's control, cannot be ignored and enhanced processing efforts might be directed to these messages as reflected in enhanced late potential amplitudes. Also, late potentials are often considered to be a "second pass analysis" that help build up conceptual representations. The observation that late potential – but not P2 - amplitudes differed between listening to

autonomy-supportive and controlling speech fits well with this account of the late potential. Moreover, we argued above that motivational prosody production relies on intentional, deliberately modulated processes. Thus, motivational expressions may be characterized by rather subtle prosodic cue variations that are prone to variability (c.f. Weinstein et al., 2014). Comparable to what has been argued for expressing vocal emotions (e.g., Paulmann & Kotz, In Press), the temporal availability of meaningful acoustic cues will also vary between motivations. Thus, motivational expressions may be inherently complex (due to their subtleness, variation and cue availability differences). This complexity has been argued to lead to more elaborated later evaluations (c.f. Wickens & Perry, 2015), and is in line with the finding that autonomy-supportive and controlling speech leads to differently modulated late potential amplitudes.

Furthermore, listeners were presented with sentences that were irrelevant for guiding their actual and immediate action. For instance, a sentence such as "*You may do this if you choose*", did not relate to any immediate action that listeners could take, given sentences were delivered out of an actual context, spoken by a stranger, and because participants had been asked to sit for ERP recordings until the end of the procedure (thus, not given a choice on how to act). As such, listeners appeared to have recognized correctly when listening to prosody conveying autonomy-support that they can disengage from this sentence as it was not self-relevant. Interestingly, *controlling* motivation continues to direct listeners' attentional resources even in a lab setting and in the absence of actionable outcomes. If true, it could be argued that controlling messages cannot be "escaped" from. This conclusion is in line with research showing that autonomy-supported individuals show more awareness of and discrimination of new information in terms of its self-relevancy (Brown & Ryan, 2003; Niemiec et al., 2010; Pennebaker & Chung, 2011; Weinstein et al., 2014), and are able to select goals that reflect truly desirable ends (Milyavskaya, Nadolny, & Koestner, 2014),

while controlling motivational climates lead to more rumination (Thomsen, Tønnesvang, Schnieber, & Olesen, 2011), more compulsion and behavioral dysregulation (e.g., Boone, Vansteenkiste, Soenens, & Van der Kaap-Deeder, & Verstuyf, 2014; Vallerand et al., 2003), and poor discrimination in decision-making (Di Domenico, Le, Liu, Ayaz, & Fournier, 2016). These findings inform extant research, suggesting that this absence of discrimination and the compulsive qualities of control may be reflected in rapid brain processing, and may be driven by basic cues toward control such as a controlling tone of voice. Importantly, the motivational literature examines the role of both motivational contexts (as was studied here) and individual differences in motivational orientations (Ryan & Deci, 2000), and future work may explore the interplay of these two constructs to build on the present findings. Similarly, because these may modulate responses to speech signals, future work may examine the role of personality (e.g., social orientation; Schirmer et al., 2008) and psychopathology (Kan et al., 2004).

Mapping of Motivational Communications

The current study set out to explore the on-line processing underlying motivational communication. The high temporal resolution of ERPs makes them an ideal methodology to investigate vocally expressed motivations in real time. However, in addition to providing information about the time-course associated with motivational prosody and speech, ERP effect distributions can also be useful in determining how motivations are processed from vocal stimuli. While the majority of ERP effects observed here were globally distributed and did not interact with our topographical factors included in the analysis, a few interactions between the condition and topography factors give rise to potentially important processing differences between motivations. First, it was found that the distribution of the P2 and late potential effects differed within the prosody-only condition, such that controlling prosody elicited larger P2 amplitudes than autonomy-supportive or neutral prosody. This effect was

distributed across the scalp. Interestingly, the same contrasts were far more localized at the later processing stage, as significant ERP differences in response to controlling versus autonomy-supportive and neutral prosody were only found at midline electrode-sites. Similarly, P2 and late potential effects were also differently distributed in the speech condition. Here, a central-right lateralized distribution was observed for the P2 effect, while the late potential was not modulated by hemisphere. Collectively, these distribution differences support the idea of multi-step approaches of social signal processing (e.g., Frühholz et al., 2016; Kotz & Paulmann, 2011; Schirmer & Kotz, 2006). These models theorize that rapid encoding of vocal characteristics is tied to auditory cortices, saliency or relevance processing is supported by the right anterior superior temporal sulcus and superior temporal gyrus, and more fine-grained meaning evaluations are linked to inferior frontal and orbito-frontal cortex. The ERPs reported here cannot be used to confirm the exact neural source of the effects; however, looking at the effects as a whole, they further support the idea that it is likely that neural generators for early and late processing stages differ. The effects observed here also provide support to models of social prosody processing which hypothesize that different brain structures mediate social signal processing at different points in time (e.g., Frühholz et al., 2016; Schirmer & Kotz, 2006).

Second, our findings allow us to speculate that processing of the two different motivational qualities autonomy-support and control relies at least partly on differing neural mechanisms. Although ERPs have a low spatial resolution, scalp distribution differences can highlight that the activity measured on the scalp has likely been generated by different neural populations (Otten & Rugg, 2005). Specifically, we found that processing of controlling information showed strongest P2 effects at left-central electrode-sites, whereas processing of autonomy-supportive information led to the most positive P2 effects at right-central electrode-sites. The difference in ERP effect distribution for the two different motivational

climates goes well with imaging data showing a differently activated brain network for behavior that is self-determined (undertaken for more autonomous reasons; Lee and Reeve, 2012). Thus, the data lend support to the idea that the two different motivation types explored here are psychological processes modulated by partly different brain networks. This speculation needs to be directly tested in future imaging studies.

Conclusions

The present study set out to shed light on the time-course underlying vocal motivational communication. Taken together, the effects observed support the idea that motivational qualities are rapidly, that is within 200 ms of sentence onset, assessed. Specifically, it seems as if the early detection of controlling prosody is used to "tag" the incoming sound as "important" or "motivationally relevant", leading to more comprehensive evaluation of the stimulus at a slightly later point in time. Crucially, this "tagging" process seems to be particularly triggered by a controlling tone of voice, but not by autonomy-supportive prosody, suggesting that the latter form of expression lacks the saliency needed to engage this early flagging process. In contrast, if motivational intentions are communicated through words and prosody at the same time, preferential processing can also be observed for autonomy-supportive speech, indicating listeners may find supportive communications consequential when they contain meaningful content. In fact, results suggest that if the information provided is salient enough, so conveyed either through a unique acoustic imprint or multiple channels, it does not matter which motivational quality the speaker is trying to convey. Hence, if formulated strongly enough, a "call to action" receives immediate attention, potentially leading to preferential processing. Once identified, the intended motivational message is continuously monitored and evaluated. In the case of non-pressuring, autonomy-supportive expressions, the stimulus is dismissed, perhaps because our design

meant these phrases were irrelevant for the listener; the same, however, was not true for controlling communications which cannot as easily be ignored.

Acknowledgement

This work was supported by a Leverhulme Trust research project grant (RPG-2013-169) awarded to Netta Weinstein and Silke Paulmann.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, *70*(3), 614-636.

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

Boersma, P., Weenink, D., 2013. Praat: doing phonetics by computer (Version 4.3.14) [Computer program]. Retrieved from http://www.praat.org/.

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

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology*, *84*(4), 822-848.

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., ... & Ryan, R. M. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, *39(2)*, 216-236.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109-134.

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). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11(4),* 227-268.

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian psychology/ Psychologie canadienne*, *49(3)*, 182-185.

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory in health care and its relations to motivational interviewing: a few comments. *International Journal of Behavioral Nutrition and Physical Activity*, *9*, 1-6.

Deci, E. L., Ryan, R. M., & Guay, F. (2013). Self-determination theory and actualization of human potential. In D. McInerney, H. Marsh, R. Craven, & F. Guay (Eds.), *Theory driving research: New wave perspectives on self processes and human development* (pp. 109-133). Charlotte, NC: Information Age Press.

Deci, E. L., Spiegel, N. H., Ryan, R. M., Koestner, R., & Kauffman, M. (1982). Effects of performance standards on teaching styles: Behavior of controlling teachers. *Journal of Educational Psychology*, *74*(6), 852 - 859.

Di Domenico, S. I., Le, A., Liu, Y., Ayaz, H., & Fournier, M. A. (2014). Basic psychological needs and neurophysiological responsiveness to decisional conflict: an event-related potential study of integrative self processes. *Cognitive, Affective, & Behavioral Neuroscience, 16(5),* 848-865.

Frühholz, S., Trost, W., & Kotz, S. A. (2016). The sound of emotions—Towards a unifying neural network perspective of affective sound processing. *Neuroscience & Biobehavioral Reviews, 68,* 96-110.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*(4), 331-362.

Grolnick, W. S., & Seal, K. (2008). *Pressured parents, stressed-out kids: Dealing with competition while raising a successful child*. Prometheus Books.

Hodgins, H. S., Brown, A. B., & Carver, B. (2007). Autonomy and control motivation and self-esteem. *Self and Identity*, *6*(2-3), 189-208.

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *Neuroimage*, *58*(2), 665-674.

Jiang, X., & Pell, M. D. (2015). On how the brain decodes vocal cues about speaker confidence. *Cortex, 66,* 9-34.

Kan, Y., Mimura, M., Kamijima, K., & Kawamura, M. (2004). Recognition of emotion from moving facial and prosodic stimuli in depressed patients. *Journal of Neurology, Neurosurgery & Psychiatry, 75,* 1667-1671.

Kanske, P., Plitschka, J., & Kotz, S. A. (2011). Attentional orienting towards emotion: P2 and N400 ERP effects. *Neuropsychologia, 49(11),* 3121-3129.

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.

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326-338.

Lee, W., & Reeve, J. (2013). Self-determined, but not non-self-determined, motivation predicts activations in the anterior insular cortex: an fMRI study of personal agency. *Social Cognitive and Affective Neuroscience*, nss029.

Levesque, C., & Pelletier, L. G. (2003). On the investigation of primed and chronic autonomous and heteronomous motivational orientations. *Personality and Social Psychology Bulletin*, *29*(12), 1570-1584.

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.

McClelland, D.C. (1987). *Human Motivation.* CUP Archive.

Milyavskaya, M., Nadolny, D., & Koestner, R. (2014). Where do self-concordant goals come from? The role of domain-specific psychological need satisfaction. *Personality and Social Psychology Bulletin*, *40*(6), 700-711.

Mitchell, R. L., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience & Biobehavioral Reviews, 37*(3), 471-479.

Mitchell, R. L., & Ross, E. D. (2013). Attitudinal prosody: What we know and directions for future study. *Neuroscience & Biobehavioral Reviews*, *37*(3), 471-479.

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.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: an integrative review of ERP findings. *Biological psychology*, *77*(3), 247-265.

Otten, L. J., & Rugg, M. D. (2005). *Interpreting event-related brain potentials*. In Todd C. Handy (Editor). *Event-related potentials: A methods handbook*. MIT Press.

Paulmann, S. & Kotz, S.A. (In Press). *The electrophysiology and time-course of processing vocal emotion expressions*. To appear in P. Belin and S. Frühholz. (Eds). Oxford Handbook of Voice Perception. Oxford University Press: Oxford.

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

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

Paulmann, S., Bleichner, M., Kotz, S.A. (2013). Valence, arousal, and task effects in emotional prosody processing. *Frontiers in Psychology* 4:345. doi: 10.3389/fpsyg.2013.00345.

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., Ott, D.V.M., Kotz, S.A. (2011). Emotional Speech Perception Unfolding in Time: The Role of the Basal Ganglia. *PLoS ONE,* 6, e17694.

Paulmann, S., Titone, D., Pell, M.D. (2012). How emotional prosody guides your way: evidence from eye movements. *Speech Communication*, 54, 92-107**.**

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.*

Pennebaker, J. W., & Chung, C. K. (2011). Expressive writing: Connections to physical and mental health. *Oxford Handbook of Health Psychology*, 417-437.

Radel, R., Sarrazin, P., & 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*, 657-674.

Radel, R., Sarrazin, P., & Pelletier, L. G. (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*, 651-674.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). 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. D. (2014). Neural correlates of inferring speaker sincerity from white lies: An event-related potential source localization study. *Brain Research, 1565*, 48-62.

Roth, G., Assor, A., Niemiec, C. P., Ryan, R. M., & Deci, E. L. (2009). The emotional and academic consequences of parental conditional regard: comparing conditional positive regard, conditional negative regard, and autonomy support as parenting practices. *Developmental Psychology*, *45*(4), 1119.

Ryan, R. M. (2012). Motivation and the organization of human behavior: Three reasons for the reemergence of a field. *The Oxford Handbook of Human Motivation*, 3-10.

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.

Schacht, A., & Sommer, W. (2009). Emotions in word and face processing: early and late cortical responses. *Brain and Cognition, 69(3),* 538-550.

Schindler, S., & Kissler, J. (2016). People matter: perceived sender identity modulates cerebral processing of socio-emotional language feedback. *NeuroImage, 134,* 160-169.

Schirmer, A. & Kotz, A.S. (2006). Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences, 10(1),* 24-30.

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

Schirmer, A., Escoffier, N., Zysset, S., Koester, D., Striano, T., & Friederici, A. D. (2008). When vocal processing gets emotional: on the role of social orientation in relevance detection by the human amygdala. *Neuroimage*, *40(3),* 1402-1410.

Silva, M. N., Marques, M., & Teixeira, P. J. (2014). Testing theory in practice: The example of self-determination theory-based interventions. *The European Health Psychologist*, *16*, 171-180.

Thomsen, D. K., Tønnesvang, J., Schnieber, A., & Olesen, M. H. (2011). Do people ruminate because they haven't digested their goals? The relations of rumination and reflection to goal internalization and ambivalence. *Motivation and Emotion*, *35*(2), 105-117.

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Léonard, M., ... & Marsolais, J. (2003). Les passions de l'ame: on obsessive and harmonious passion. *Journal of Personality and Social Psychology*, *85*(4), 756.

Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration, 23(3),* 263-280.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, *92*(4), 548.

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., & 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., Hodgins, H. S., & Ryan, R. M. (2010). Autonomy and control in dyads: Effects on interaction quality and joint creative performance. *Personality and Social Psychology Bulletin*, 0146167210386385.

Weinstein, N., Przybylski, A. K., & Ryan, R. M. (2013). The integrative process new research and future directions. *Current Directions in Psychological Science, 22(1),* 69-74.

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

Wickens, S., & Perry, C. (2015). What Do You Mean by That?! An Electrophysiological Study of Emotional and Attitudinal Prosody. *PloS one,10*(7), e0132947.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research, 156,*249–268.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T. &

Weinstein, N. (2016, January). Motivating prosody alone can change social and emotional experiences. In N. Weinstein (Chair), *The Social Voice*. Symposium conducted at the meeting of the Society for Personality and Social Psychology, San Diego, US.

Figure Captions:

Figure 1

Spectrograms

The illustration shows example spectrograms and waveforms for stimuli used. Panel A shows examples for the prosody condition ("You often read books at night" spoken in neutral, controlling, or autonomy-supportive prosody), while Panel B shows examples for the speech condition ("You often read books at night" spoken in neutral, "You better do it my way" spoken in controlling, and "You may do this if you choose" spoken in autonomy-supportive prosody). Spectrograms show visible pitch contours and were created with Praat (Boersma & Weenink, 2013).

Figure 2

P200 and Late Component Effects

Figure 2 (A, B, C, D). The illustration shows event-related brain potentials in response to motivational stimuli at selected electrode-sites from 100 ms before the start of the sentence up to 800 ms into the sentence. Panel A displays effect for the prosody condition, panel B for the speech condition, and panels C & D compare these effects for autonomy-supportive (C) and controlling speech (D). Negativity is plotted upwards.

Figure 3

Topography

Figure 3 (A, B, C, D). The illustration shows topographical maps for the P2 and late component time windows illustrating the distribution of the responses to neutral, autonomy-supportive and controlling prosody (A, C) as well as neutral, autonomy-supportive and controlling speech (B, D).