Now Listen to This! Evidence from a Cross-Spliced Experimental Design Contrasting Pressuring and Supportive Communications

Silke Paulmann[1]* Netta Weinstein[2] and Konstantina Zougkou[1]

[1] University of Essex

[2] Cardiff University

***Corresponding Author Information**

Silke Paulmann
University of Essex
Department of Psychology and Centre for Brain Science
Wivenhoe Park
Colchester, CO4 3SQ, UK
Email: paulmann@essex.ac.uk
Phone: +44-(0)1206-873422
Fax: +44-(0)1206-873801

Abstract

Motivating communications are a frequent experience within daily life. Recently, it has been found that two types of motivations are spoken with distinct tones of voices: control (pressure) is spoken with a low pitched, loud tone of voice, fast speech rate, and harsh sounding voice quality; autonomy (support) is spoken with a higher pitched, quieter tone of voice and a slower speech rate. These two motivational tones of voice also differentially impact listeners' well-being. Yet, little is known about the brain mechanisms linked to motivational communications. Here, participants were asked to listen to semantically identical sentences spoken in controlling, neutral, or autonomy-supportive prosody. We also presented cross-spliced versions of these sentences for maximum control over information presentation across time. Findings showed listeners quickly detected whether a speaker was providing support, being pressuring, or not using motivating tones at all. Also, listeners who are pressured do not seem to respond anew when a supportive motivational context arises, but those who had been supported are affected by a newly pressuring environment. Findings are discussed in light of motivational and prosody literatures, and in terms of significance for the role of motivational communications on behavior.

Keywords: motivational prosody; self-determination theory; tone of voice; ERPs;

Word count: 7327 (excl. references)

1. Introduction

Many daily interactions are focused on attempts to motivate others – to energize them to action. For example, a familiar message among collaborating researchers might be similar to "time to get this [paper] out!", which, depending on the tone of voice used by the speaker, could be interpreted in several ways. Past research shows that when spoken with a relatively low pitched, but loud tone of voice, as well as fast speech rate and harsh sounding voice quality, the listener will be driven to action because they are attempting to conform to imposed expectations (Weinstein, Zougkou, & Paulmann, 2018). In contrast, listening to the same message spoken in a higher pitched, quieter tone of voice and a slower speech rate may lead the listener to feel that the action (in this example, to complete a writing task) may be linked to their own interests, beliefs or values (e.g., that writing is a good idea because the author would like to communicate her or his useful findings). This latter style of communicating gives the listener room to remember that the task might be personally valued or interesting.

The role of such motivational language has long been the focus of self-determination theory (SDT; Deci & Ryan, 1985; Ryan & Deci, 2000), a theoretical framework that studies human motivation. SDT describes the two motivational qualities alluded to in our examples as "controlling" (here, motivation comes from the pressuring tone of voice) and "autonomy-supportive" (the listener is left with a sense of choice and volition) and argues that these qualities are routinely used to attempt to change others' behavior. Although still in its infancy, research aiming to specify the neural underpinnings of those two motivational qualities has started to emerge. For instance, Lee and Reeve (2013) asked participants to read about and imagine themselves in situations in which activities were either self-endorsed (e.g., writing an enjoyable paper) or not self-endorsed (e.g., writing a paper for extra course credit). Imaging results revealed higher anterior insular cortex (AIC) activation for self- versus non self-determined activities. In a follow-up study, the same authors (2017) also reported sub-cortical activations in the striatum for self-endorsed activities, and, crucially, functional interactions between the AIC and striatum. These reports support the notion that the motivational significance of a stimulus might be signaled from a range of sub-cortical nodes such as the amygdala, nucleus accumbens, substantia nigra, and ventral tegmental area to the AI. In fact, it has been suggested that the so-called "salience network" might be playing a dominant role when stimuli are particularly motivationally relevant (see DiDomenico and Ryan, 2017). So far, this neural evidence only addresses motivational messages communicated predominantly through *words* (e.g. studying *for fun* vs. *for a grade)* and it remains to be seen if comparable results are elicited when motivations are communicated through tone of voice. While research such as this has extensively explored which words may be used to communicate motivations (Hodgins, Brown, & Carver, 2007; Levesque & Pelletier, 2003; Radel, Sarrazin, & Pelletier, 2009; Ryan, 2012; Weinstein & Hodgins, 2009), relatively little is known about how the prosody, or tone of voice, used to deliver motivational messages (Weinstein, Zougkou & Paulmann, 2014; Weinstein et al., 2018, Zougkou, Weinstein, Paulmann, 2017), elicits differential responses from listeners.

Thus, the current study set out to further investigate the on-line processing of motivational prosody, contrasting controlling and autonomy-supportive motivational communications in a new way to enhance our understanding. In particular, the current study paired experimental manipulations of motivational tones with a cross-splicing design to maximize control over when and how motivational information would be delivered to listeners and for the first time account for potential confounds by the speed by which information is delivered. Further, cross-splicing allowed us to examine immediate reactions to motivational tone as listeners adjusted to violations in their expectations that a motivational sentence would be either supportive or controlling. Second, this study was first to examine how listeners respond to motivational content to which they actively attend, increasing the generalizability of laboratory findings to real-life interactions between motivators and listeners where listeners are not passive recipients of content. In sum, this study represented an effort for a first robust test of the underexplored neural mechanisms underlying the perception of social prosody to help illuminate how listeners respond to motivational messages as conveyed through prosody in real-time.

1.1 Motivational Prosody

Studies in this area have found support for the view that motivations are expressed through distinct prosodic patterns. For example, in the first study on this topic, university students were asked to read out sentences that contained words that were specific to either an autonomy-supportive or controlling motivational framing (e.g., "you have to do this my way" *or* "you are free to do this"). Results revealed that when asked to project tone onto these sentences, students selected to express autonomy-supportive communications with a quieter voice, slower speech rate, and milder sounding voice quality than sentences linking to controlling environments than when expressing controlling communications (Weinstein et al., 2018). This was taken as first evidence that speakers who were simply asked to read out sentences as if they really meant them would use different vocal patterns for each of the two motivation types. In later research, similar patterns for prosodic indicators were found when actors intoned sentences that either contained or lacked motivational biasing words.

The latter set of data were particularly intriguing; even when reading motivationally neutral terms (e.g., "time to leave"), but with a motivational mindset that was either controlling or supportive, speakers expressed control with a low pitched, loud, and harsh voice using a fast speech rate, while autonomy-supportive prosody was conveyed using a high pitched, less loud voice while slowing down speech rate (Weinstein et al., 2014; 2018). Similar results were observed in Dutch speaking parent-child interactions, where acoustic analyses supported previous results in a different language group and in naturalistic observations (Paulmann, Vrijders, Weinstein & Vansteenkiste, 2018). The finding that different motivational intentions can be expressed through prosodic cues alone is in line with the large body of literature showing that *emotional* and *attitudinal* intentions can be distinguished based on acoustic cues alone (e.g., Banse & Scherer, 1996; Mitchell & Ross, 2013).

1.2. Motivational Prosody: Evidence from electrophysiology

Accumulating evidence suggests that speakers use specific acoustic cues to express motivational intentions. Moreover, listeners *respond* as a function of the motivational tone used, demonstrating higher well-being and more prosocial behavior intention after listening to autonomy-supportive tones (Weinstein et al., 2018). More immediate to the present research is previous work using event-related brain potentials (ERPs) that examined and delineated the time-course of the neural processing mechanisms underlying motivational prosody perception (Zougkou etal., 2017). In this study, while engaged in an implicit motivation processing task, i.e. a task which did not require participants to focus on the motivational properties of the stimuli, passive listeners were presented with materials spoken in autonomy-supportive, controlling, or neutral tone of voice. Sentence context either matched the motivational qualities of the prosody used (e.g., "You have to do it my way" spoken in a controlling tone of voice), or contained no motivationally biasing words (e.g., "Why don't we meet tomorrow" spoken in a controlling voice). Results showed that listening to motivations conveyed through tone of voice alone led to enhanced P2 amplitudes in response to controlling as opposed to neutral messages, though no differences were found for autonomy-supportive tones, suggesting that listeners are particularly receptive to controlling messages. In addition, enhanced long positive potential (LPP) amplitudes were observed in response to controlling when compared to neutral prosody. Taken together, results suggest that listeners respond rapidly to controlling prosody and continue to monitor speech that is spoken in a pressuring tone of voice, even when not explicitly focusing on motivational qualities of stimuli (Zougkou et al., 2017). Yet, there is little understanding of whether the preferential processing mechanisms observed for controlling prosody will be upheld even when listeners are asked to evaluate, that is *explicitly* process tone of voice. However, exploring this further is important because in actual interactions between motivators and listeners, it is hardly the case that motivated individuals are passively listening (they may do so, for example, when hearing another child in the same classroom receive a motivating speech from a teacher, or by overhearing one's father and mother interacting); in most cases, the listener is to some degree engaged in the conversation through which motivation is delivered.

1.3. When incoming information is unexpected: prosodic expectancy violations

In addition to exploring different processing stages of real-time prosody perception of emotional, attitudinal, or motivational information (see e.g., Paulmann & Kotz, 2015, In Press; or Mitchell & Ross, 2013 for reviews), some have employed cross-splicing designs to investigate whether prosodic functions are processed differentially, while controlling for potential confounding effects which may result from differential, but unintended, delay in presentation of meaningful acoustic information across conditions (e.g., Paulmann, Jessen, Kotz, 2012). For instance, to explore whether linguistic and emotional prosody run a similar time-course, Paulmann and colleagues (2012) applied a cross-splicing paradigm that allows to temporally control when prosodic information becomes available to the listener (c.f. Kotz & Paulmann, 2007). In their study, they merged a start of a sentence with an ending that deviated from either the emotional or linguistic prosodic expectancy of the listener. For instance, listeners heard (in German) the start of the sentence "He has" spoken in a neutral prosody, followed by "watered the plants and cut them" spoken in an angry tone of voice or spoken as a question. Results showed that emotional and linguistic prosodic expectancy violations were met with a prosodic expectancy positivity (PEP) which differed both in terms of its speed (occurring earlier or later) and its location or distribution across the head (Paulmann et al., 2012). These findings suggested a) that emotional and linguistic prosody follow a different time-course and b) that emotional prosody is processed more quickly. The ERP distribution differences also suggest that partially different neural mechanisms are at play for the two different prosodic functions. Thus, the cross-splicing methodology has proven useful in delineating neural mechanisms underlying real-time prosody processing.

In fact, many other studies have adapted this cross-splicing paradigm when attempting to differentiate prosody functions (e.g., linguistic vs. emotional; attitudinal vs. emotional) or different prosody types (e.g., anger vs. happiness). For instance, Kotz and Paulmann (2007) studied the comparative nature of emotional prosodic and semantic information. Results revealed that combined semantic and prosodic expectancy violations elicited a negative ERP with an earlier onset than the PEP reported in response to emotional prosodic only expectancy violations. This pattern not only suggests that listeners rapidly integrate verbal and non-verbal information, it also showed that the nature of the information available (e.g., semantics and/or prosody) seems to determine how quickly listeners detect deviances from what is expected. Indeed, detection latency and distribution varies across and within emotional (Chen Zhao, Jiang & Yang, 2011; Kotz & Paulmann, 2007; Paulmann et al., 2012; Paulmann & Kotz, 2008a; Paulmann, Pell & Kotz, 2008), linguistic (Astésano, Besson & Alter, 2004; Eckstein & Friederici, 2005; Paulmann et al., 2012), and attitudinal expectation violation paradigms (Wickens & Perry, 2015). Speaker, task, or emotional category differences can also modulate PEP latencies (e.g., Kotz & Paulmann, 2007; Paulmann, Ott, Kotz, 2011) and the component has been reported for both tonal (e.g., Mandarin; Chen et al., 2011) as well as non-tonal languages (e.g., German and French; Kotz & Paulmann, 2007; Paulmann & Kotz, 2008a; Paulmann et al., 2012; Astésano et al., 2004). To summarize, a large body of evidence suggest that the cross-splicing paradigm is crucial for clearly determining whether types of prosody follow the same time-course and whether some information is more relevant to the listener, and so is processed in a favorable manner.

1.4. The present study

To investigate further how the two motivational qualities of interest here, namely autonomy-support and control, are processed in real-time and which information type might take precedence during on-line processing, we thus utilized stimuli from Zougkou et al. (2017) presenting either autonomy supportive, neutral, or controlling motivating tones, and expected to conceptually replicate findings that controlling motivational tones would be preferentially attended to as compared to neutral tones (Hypothesis 1). In this study, we also anticipated to see a preferential effect for autonomy support over neutral sentences given participants actively attended to the task (Hypothesis 2). Moreover, assuming that motivational stimuli activate the salience network (c.f. Domenico and Ryan, 2017), we also hypothesized that controlling sounding materials will most captivate listeners' attention (Zougkou et al., 2017) and will thus lead to most pronounced ERP amplitudes. In addition, we applied the prosodic expectancy violation paradigm. Specifically, this procedure allows us to determine if deviance detection for the two motivation qualities differs and which motivation type listeners may attend to preferentially. The only study evaluating the time course of these motivational communications (Zougkou et al., 2018) identified that listening to controlling tones led to more enhanced ERP components (P2, LPP) than either supportive or motivationally neutral tones. However, it was not clear to what extent this finding was due to a confounding effect because controlling tones may have provided, incidentally, more meaningful prosodic information earlier on in the sentence than did supportive tones. It was also not apparent whether listening to autonomy support, a motivational style which has also been linked to deeper processing of self-relevant content (Luyckx, Soenens, Berzonsky, Smits, Goossens, & Vansteenkiste, 2007; Weinstein, Przybylski, & Ryan, 2013), failed to show differences to listening to neutral prosody because a passive listening task was applied.

To address these two critical shortcomings, we developed two sets of cross-spliced stimuli. In the first set, a neutral sounding beginning (start of sentence) was merged to a motivational (either autonomy-supportive or controlling) sounding end. The majority of past research would suggest that deviance detection is relatively similar across emotional categories (e.g., Kotz & Paulmann, 2007; Paulmann & Kotz, 2008a; Wickens & Perry, 2013), and temporal differences between categories are only rarely observed (Paulmann et al., 2011). Given this, we hypothesized that transitions from neutral to autonomy-supportive or controlling prosody would be processed, but in a similar fashion. Thus, if true that the PEP component predominantly reflects a "function" change (e.g., neutral to emotional; statement to question prosody), listeners would detect that the speaker changed from using a neutral to motivational tone of voice, but would not prioritize attending to controlling prosody over autonomy supportive prosody (Hypothesis 3). Thus, we would expect to see no PEP modulation differences between the two spliced conditions when compared to neutral prosody (i.e. transitions from neutral to autonomy-supportive as well as transitions from neutral to controlling prosody lead to the same PEP modulation differences when compared to neutral prosodic materials).

To further explore the possibly prioritized processing of controlling versus autonomy supportive tone of voice (c.f. Zougkou et al., 2017), we developed a second set of stimuli comprising of sentences that changed from one motivational prosody quality to another (e.g., the first part of sentence was spoken in a supportive tone of voice while the second part was spoken in a pressuring tone). If true that controlling prosody is attended to more quickly than autonomy-supportive prosody, we expected an earlier onset latency for the PEP in response to transitions from autonomy support to control than when the transition occurs the other way around (Hypothesis 4).

2. Methods

2.1 Participants

Twenty-six native English speakers from the University of Essex participated in the study. We excluded participants who were left with less than 1/3 of trials in one or more conditions after data cleaning and removing incorrectly answered trials to ensure a signal-to-noise ratio which is comparable to previous studies applying this paradigm. Six of the remaining participants were male and 10 female with a mean age of 23.25 years (*SD* = 7.36, range: 18 to 49 years). All participants reported right-hand dominance assessed by an adapted version of the Edinburgh Handedness Inventory (Oldfield, 1971). None of the participants responded in the affirmative to questions assessing whether they have hearing difficulties or take medication for psychopathology or mood disorders. All participants gave informed consent before completing the study, which was ethically approved by the University of Essex Science and Health Faculty Ethics Sub-committee.

2.2 Materials

2.2.1 Sentences. All materials were spoken by two female actors. From an established inventory used in social psychological research on affective outcomes of motivational prosody (Weinstein et al., 2018), we employed 25 semantically neutral sentences for each speaker (e.g., Join me at the park later; Come to visit me next week). These were expressed in a tone of voice characterizing one of the motivational qualities (autonomy support or control) or in a neutral tone of voice (prosody condition). In addition, we included 25 sentences for which semantic content and prosodic realization matched for each motivational quality (i.e., autonomy support, control) as well as the neutral category, which were treated as filler materials given our interest in motivational prosody and ensured that participants heard a variety of sentences. Previous ratings indicated that autonomy-supportive speech (semantic and prosody matched) and autonomy-supportive prosody (neutral semantics) sentences were recognized as sounding more supportive of choice (M = 4.12, SD = 0.34 and M = 3.49, SD = 0.37) and less pressuring (M = 1.78, SD = 0.34 and M = 1.86, SD = 0.30). Similarly, controlling speech and controlling prosody were reported as sounding more pressuring (M = 4.05, SD = 0.33 and M = 3.4, SD = 0.54) and less supportive of choice (M = 1.70, SD = 0.27 and M = 2.37, SD = 0.51). The length of the stimuli in terms of number of words that each sentence contained was matched across conditions (range 3-9 words). We ran acoustical analyses for selected materials and found significant differences (all p's $\leq .001$) between motivational conditions in terms of pitch (highest for autonomy support, followed by control and neutral prosody), amplitude (highest for control, followed by autonomy support and neutral) and speech rate (slowest for neutral, followed by control and autonomy support). Means for acoustic measurements extracted with *praat* software (Boersma & Weenink, 2016) can be found in Table 1.

Spliced Sentences. In addition to presenting non-violated materials described above, two splicing conditions were created using semantically neutral sentences that differed in the prosody used when expressing them (See Table 2 for a summary of all conditions). For the *neutral spliced with motivation* condition, we thus took a neutrally intoned beginning and spliced it to a motivationally (autonomy supportive or controlling) intoned ending. This resulted in 100 (25 neutral/autonomy-supportive prosody and 25 neutral/controlling prosody x 2 speakers) cross-spliced sentences.

In a second condition (*spliced motivation*), sentences were spliced in a different manner: sentences started out with autonomy-supportive tone and ended with controlling tone, or vice versa. There were hence 50 autonomy-supportive/controlling prosody and 50 controlling/autonomy-supportive prosody cross-spliced sentences. Sentences were cross-spliced applying phonetic rules that would allow easy identification and isolation of phonemes (Kotz & Paulmann, 2007); given our sentence lengths, we decided to splice materials after the first two or three words (e.g., "*Have a look* at the paper"). On average, for the neutral spliced with motivation condition, the mean splicing point occurred ~ 530 ms after sentence onset; for the spliced motivation condition, the mean splicing point occurred ~430 ms after sentence onset for sentences containing an autonomy-supportive sounding beginning and ~ 510 ms for sentences starting with a controlling sounding prosody.

	Condition	Mean F0	Range F0	Range loudness	Duration
1	Autonomy	242.51 Hz	168.04 Hz	36.51 dB	1.66 secs
2	Control	217.07 Hz	210.32 Hz	42.60 dB	1.82 secs
3	Neutral	204.18 Hz	161.16 Hz	37.57 dB	1.48 secs
4	Neutral/autonomy	227.36 Hz	181.60 Hz	38.28 dB	1.43 secs
5	Neutral/control	207.78 Hz	169.86 Hz	39.47 dB	1.44 secs
6	Autonomy/control	235.96 Hz	216.51 Hz	37.65 dB	1.35 secs
7	Control/autonomy	239.23 Hz	180.74 Hz	43.81 dB	1.41 secs

Table 1 summarizes acoustic data extracted for each condition. Note: Hz = Hertz; dB = decibels; secs=seconds.

	Condition	Start	End
1	Autonomy	Autonomy	Autonomy
2	Control	Control	Control
3	Neutral	Neutral	Neutral
4	Neutral spliced with motivation	Neutral	Autonomy
5		Neutral	Control
6	Spliced motivation	Autonomy	Control
7		Control	Autonomy

Table 2 summarizes all conditions. The first three conditions contain non-violated materials while the last four conditions all contain spliced materials.

2.3. Procedure

Each participant was prepared for EEG recordings in a sound attenuated room. After preparation for EEG recording, participants received instructions about the experiment: they were asked to listen to sentences which would be presented one at a time and then judge if they felt the speaker supported a sense of “choice”, “pressure” or “neither”. Participants started with a practice round. All stimuli were presented with SuperLab 5 (Cedrus Corporation, San Pedro, California) in a fully randomized order. A total of 450 trials were presented with a break after every 90 trials. Participants were seated approximately 100 cm away from a computer screen and were asked to look at a fixation cross that remained in the middle of the screen and not to blink while they listened to each sentence. Three-hundred milliseconds later, a stimulus was presented via speakers located to both sides of the computer screen. The fixation cross remained on the screen during spoken sentence presentation. After the end of each sentence, participants were presented with a blank screen for 200ms before the response screen appeared. Instructions asked participants to indicate if the speaker sounded as if they supported a sense of “choice”, “pressure” or “neither”. Answers were then given using a key-pad (the button for “neither” was always in the middle, while presentation order of “choice” and “pressure” was counter-balanced across participants). The question remained on screen until the participant responded. An inter-stimulus interval (ISI) of 1500ms preceded the next stimulus presentation. The computerized task had a run-time of approximately 30 minutes.

2.4. ERP recording

Electrophysiological (EEG) data were collected using 63 Ag/AgCl electrode channels embedded on a custom-made cap (waveguard) following the International 10-20 system. Signals were recorded continuously using a 72-channel Refa amplifier (ANT) with an online band pass between DC and 102Hz, and were digitized at 512Hz sampling rate. Electrode impedances were kept below 7KΩ. Bipolar horizontal and vertical electro-oculograms (EOGs) using disposable Ambu Blue Sensor N ECG electrodes were recorded for artefact rejection purposes. The electrode on the left mastoid was used as an online reference and CZ served as a ground electrode.

Data Analysis

Data were filtered off-line with a bandpass filter set between 0.01 and 30 Hz. After re-referencing data to the average of the left and right mastoids, automatic rejections for muscle or electro-oculugram (EOG) artefacts above 30.00 µV were performed with EEProbe Software. Subsequently, all data were visually inspected to remove undetected artefacts. In total, 13% of data on correct responses was rejected (range for different conditions: 10-14%). Trials for each condition and each electrode-site averaged from 200 ms before stimulus onset to 800 ms after stimulus-onset. All data were baseline corrected by subtracting the mean voltage of the baseline window (-200 to 0 ms) from the averaged signal. Depending on condition, ERPs were time-locked to the onset of the sentence, or the onset of the splicing point.

To explore topographical distribution of effects, we followed previous research (Paulmann et al., 2013) and included *regions of interest* (ROI) in all statistical analyses: left frontal (F5, F3, FC5, FC3), right frontal (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). Midline electrode-sites (Fz, Cz, CPz, Pz) were analyzed separately.

Analyses for the prosody condition. For ERP analysis on non-violated materials, time windows of interest for mean amplitude measures were based on visual inspection and have also been the focus in past research on motivational prosody (Zougkou et al., 2018). Two time-windows were of interest: 170 – 230 *ms* (P2) and 350 – 600 *ms* (late potential) following sentence onset. We explored the time-course underlying on-line processing of motivational and neutral prosody when participants' task focuses on the speaker's tone of voice by comparing all three prosody conditions with one another.

Analyses for expectancy violations. ERP components of interest for the expectancy violations analysis were determined based on visual inspection and a 50 ms time-line analysis. Two separate comparisons were of interest: 1) Transitions from neutral to motivational prosody (i.e., autonomy-supportive prosody vs neutral beginning spliced to autonomy-supportive prosody end; and controlling prosody vs. neutral beginning-controlling prosody end). 2) Looking at transitions from one motivational prosody to another (i.e. autonomy-supportive prosody vs. controlling prosody spliced to autonomy-supportive prosody; and controlling prosody vs. autonomy-supportive prosody spliced to controlling prosody). Two separate time-windows were of interest: 100 - 250 ms and 400 - 700 ms post splicing point onset.

ERP Results

Prosody Comparison

To test Hypotheses 1 and 2, namely that motivational and neutral prosody will be differentiated in an early and late time window (Hypothesis 1) and that controlling and autonomy-supportive prosody can be differentiated (Hypothesis 2), mean ERP amplitudes were analyzed with a general linear model (GLM) treating *prosody* (autonomy support, control, neutral), and *region of interest (ROI)* (left/right frontal, left/right central, left/right posterior and midline) as repeated-measures factors. Main effects and interactions that reached significance ($p < .05$) were followed by simple main effects. Greenhouse-Geisser correction was applied to all measures with greater than one degree of freedom in the numerator. Figure 1 displays ERP waveforms in response to motivational prosody.

P2 (170-230 ms)

In a first analysis, a main effect of *prosody* emerged, $F(2, 30) = 6.52$, $p < .01$, suggesting that the motivationally laden and neutral tones of voice overall elicited differently modulated P200 components. Planned post-hoc contrasts revealed significant P200 differences between controlling and neutral prosody, $F(1, 15) = 10.55$, $p < .001$, showing enhanced P200 components for controlling prosody. The contrasts between autonomy-supportive and neutral, $F(1, 15) = 3.89$, $p = .07$, and autonomy-supportive and controlling prosody, $F(1, 15) = 3.56$, $p = .08$, failed to reach significance. The main effect was qualified by a significant interaction between *prosody* and *ROI,* $F(12, 180) = 4.53$, $p < .01$, suggesting that some prosody effects were most pronounced at certain electrode-sites. Follow-up comparisons by *ROI* showed differences in mean P200 amplitudes for all of the comparisons. The patterns observed are summarized in Table 2; for all comparisons, controlling prosody elicited the strongest P200, followed by autonomy support, and finally neutral prosody. In short, analyses for the P2 component revealed a fronto-centrally distributed P200 component which distinguished the different motivational and neutral prosodies, supporting Hypotheses 1 and 2.

Late Potential (350-600 ms)

Further, as predicted by Hypotheses 1 and 2, there was a main effect of *prosody* in the late potential, $F(2, 30) = 5.43$, $p = .01$, suggesting differently modulated amplitudes for the three different prosodies. A non-significant interaction between *prosody* X *ROI* interaction, $F(12, 180) = 2.05$, $p = .09$, was followed-up by planned post-hoc contrasts for each *ROI*. Table 2 summarizes data that showed significant differences between listening to neutral versus controlling prosody at all ROIs. In addition, autonomy-supportive and controlling prosody elicited different amplitudes at left and right posterior electrode-sites while the contrast between autonomy support and neutral led to differences at left frontal sites.

Time Window	Contrast	LF	LC	LP	ML	RF	RC	RP
P2	A-S vs. NE	4.50			4.06	6.07	5.14	
		.051	.	.	*.063*	*.026*	*.039*	.
	CO vs. NE	13.32	*7.46*		13.22	14.14	9.38	4.38
		.001	*.016*	.	*.002*	*.002*	*.007*	*.054*
	A-S vs. CO	3.55	.032		6.18			
		.079	*5.59*	.	*.025*	.	.	.
Late potential	A-S vs. NE	6.92				3.84		
		.019	.	.	.	*.070*	.	.
	CO vs. NE	8.53	5.63	6.16	12.95	10.07	9.54.	12.01
		.011	*.031*	*.025*	*.003*	*.006*	*.008*	*.004*
	A-S vs. CO			4.94				3.62
		.	.	*.042*	.	.	.	*.076*

Table 2 summarizes post-hoc contrasts for P200 and late potential components at all ROIs. Top lines in each cell refer to F-values, bottom lines to *p*-values (if < or = .08; else no data are reported). A-S refers to the autonomy-support prosody condition; NE refers to the neutral prosody comparison condition; CO refers to controlling prosody. LF = left frontal electrode sites; LC = left central; LP = left parietal; ML = midline; RF = right frontal; RC = right central; RP = right parietal.

Expectancy Violations

Hypotheses 3 and 4 aimed to test in how far listeners can detect changes in speaker tone and whether this change detection will depend on motivational quality used. For each expectancy violation condition (i.e. transitions from neutral to motivational, or transitions from motivational to opposite motivational prosody), mean amplitudes were analyzed with a general linear model (GLM) treating *violation* (spliced, non-spliced materials), *prosody* (autonomy-support, control), and *region of interest (ROI)* (left/right frontal, left/right central, left/right posterior and midline) as repeated-measures factors. Main effects and interactions that reached significance ($p < .05$) were followed by simple main effects. Greenhouse-Geisser correction was applied to all measures with greater than one degree of freedom in the numerator.

Early time-window (100-250 ms)

Transitions from neutral to motivational prosody. There was no main effect of *violation*, $F(1, 15) = 2.04$, $p = .18$, but the main effect of *prosody* turned out to be significant, $F(1, 15) = 9.01$, $p < .01$, showing that amplitudes in response to listening to autonomy-supportive prosody were more negative-going than amplitudes in response to controlling prosody. No other effects reached significance in this early time-window, suggesting that changes from a neutral to a motivational tone of voice were not detected at this early stage.

Motivational transitions. This cross-splicing condition specifically tested Hypothesis 4 which predicted that transitions to controlling prosody might be preferentially processed compared to transitions to autonomy support prosody. There was no main effect of *violation*, $F(1, 15) = 1.39$, $p = .26$, and the main effect of *prosody* was non-significant, $F(1, 15) = 3.65$, $p = .08$, but an interaction between *violation* X *prosody* was present, $F(1, 15) = 19.76$, $p < .001$. For sentences expressing autonomy support, no significant *violation* effect was found, $F(1, 15) = 2.64$, $p = .13$, that is no differences emerged between sentences expressing autonomy support compared to those that started with a controlling and ended with an autonomy-supportive voice; however, listening to sentences which started out sounding autonomy supportive but ended with controlling prosody led to an enhanced early negativity when compared to sentences that expressed control from sentence beginning, $F(1, 15) = 16.62$, $p = .001$. This finding suggested that listeners rapidly process transitions from autonomy-supportive into controlling prosody but not vice versa.

Late time-window (400-700 ms)

Transitions from neutral to motivational prosody. As predicted in Hypothesis 3, there was a main effect of *violation*, $F(1, 15) = 49.63$, $p < .0001$, revealing more positive-going amplitudes for violated as opposed to non-violated materials. The main effect of *prosody* also reached significance, $F(1, 15) = 10.26$, $p < .01$. ERPs in response to autonomy-supportive prosody were more negative than amplitudes in response to controlling prosody. In addition, there was a significant *violation* X *ROI* interaction, $F(6, 90) = 8.44$, $p < .001$. Follow-up analyses by *ROI* revealed that the positivity in response to spliced sentences was broadly distributed in all regions of interest (ps $\leq .01$).

Motivational transitions. Further, again as predicted in Hypothesis 3, a main effect of *violation* was present, $F(1, 15) = 5.45$, $p < .05$, showing a positivity in response to spliced sentences irrespective of whether the transition was from autonomy support to control or the other way around. No other effects reached significance. This main effect demonstrates that listeners can detect changes in speaker tones at this point in time, irrespective of which tone the speaker changes to.

Discussion

Although many studies have examined the ways in which emotions are communicated through tone of voice and processed by listeners, we have little understanding of how *motivating* tones of voice – which are used to energize others to action – are processed, despite their ubiquity, and importance for listener well-being and behavior (Ryan & Deci, 2017). The current study was first to explore the temporal processing underlying motivational qualities, autonomy support and control, in a systematic and robust manner. We specifically focused on two questions of interest: First, we investigated the comparative processing of motivational prosody using a cross-splicing technique to specify if the two motivational qualities of interest run a similar time-course; this directly tested the extent these motivating tones are meaningfully different to listeners. Second, we tested whether previously observed preferential processing of salient motivation-revealing auditory information (Zougkou et al., 2017) occurs even when listeners are asked to focus on the prosody used by the speakers. Results, in sum, revealed that both motivational intentions, i.e. autonomy-supportive and controlling prosody, are differentiated from each other as well as from neutral prosody within 200 ms after sentence onset. Thus, listeners processed the two qualities of motivational tones in meaningfully different ways.

Paying attention to motivational cues early on

Building on previous work, we found that two vocal motivations, autonomy support (receptive and encouraging), and control (pressuring), are differentiated from non-motivational vocal messages as early as 200 ms after sentence onset and continue to be processed differently at later time-points. This finding is important because the presence of a more pronounced early P2 component has been argued to signal that a stimulus is more relevant to listeners (e.g., Paulmann et al., 2013; Schirmer et al., 2013); in the context of this study then, it appeared that both autonomy-supportive and controlling tones of voice were perceived to be more relevant to listeners than the same sentences spoken in neutral tones. In other words, given these results we might speculate that motivating tones of voice, even when independent from the better-studied effects of motivating words (e.g., DeMuynck et al., 2017; Hodgins, Brown, & Carver, 2007; Levesque & Pelletier, 2003; Radel, Sarrazin, & Pelletier, 2009), are in themselves meaningful to listeners. The results thus support the notion that incoming auditory information is scanned for saliency before it is passed on to higher-order cognitive processes (e.g., Schirmer & Kotz, 2006; Kotz & Paulmann, 2011). The P2 component is known to be sensitive to pitch contour variations (e.g., Friedrich, Alter, Kotz, 2001) and loudness (e.g., Picton et al., 1970), but has also been linked to signal spectral complexity in instrumental tones (Shahin et al., 2005). In line with this, Spreckelmeyer and colleagues (2009) report enhanced amplitudes for stimuli that contained more high frequency information. We previously showed that controlling sounding materials are expressed with more intensity in high frequency bands (Weinstein et al., 2018), suggesting that processing of this frequency information was also at play here where we observe most pronounced P2 amplitudes for controlling sounding prosody. In short, we propose that the P2 response elicited by motivational prosody can be tied to processing a combination of acoustic cues linked to the perception of pitch, loudness, timbre or other voice quality features. Collectively, these cues signal the motivational relevance of stimuli.

The current findings partially contradict previous ones examining motivational tones when listeners were not specifically asked to pay attention to speakers' vocal cues. In previous research, neutral and controlling tones were processed differently, while neutral and autonomy-supportive prosody were processed in a similar fashion (Zougkou et al., 2017). The lack of early differentiation between neutral and autonomy-supportive prosody was linked to the latter possibly missing salient enough *prosodic* cues as differentiation between neutral and autonomy support was found in the same study for messages that also contained biasing motivational content (e.g., "you may [do this, if you choose]" spoken in an autonomy-supportive voice). Yet the reliance on *passive* listeners of motivational content was an important limitation. Specifically, passive listening may lead to inhibited effects of autonomy supportive tones, since support – in general – is a gentler form of motivation. It is more reliant on listener engagement, in part because it motivates through inspiring self-relevant processing of motivational content (Luyckx et al., 2007; Weinstein et al., 2013).

Thus, in this study, we asked listeners to pay attention to the voice cues used by the speakers and found that neutral and autonomy-supportive prosody now also differentiate quickly after a message starts to unfold. It seems that listeners evaluate the incoming information more thoroughly when task focus is specifically directed to a speaker's prosody and thus also notice more subtle acoustic cue use differences.

Crucially, the data suggest that messages communicating the two theoretically opposing motivational qualities are rapidly assessed by listeners. This assessment presumably allows listeners of motivational communications to prepare for undertaking appropriate responses, arguably in these cases, to follow the directive requests. In particular, it has been argued that this initial evaluation of prosodic information leads to increased P2 amplitudes, that is messages that contain action-relevant cues conveyed through prosody are specifically attended to (c.f., Jessen & Kotz, 2011; Paulmann et al., 2013). This hypothesis goes well in line with P2 amplitude modulations observed here: listening to controlling prosody elicited strongest P2 amplitudes in listeners, that is, the motivational quality most often used to create an instant reaction or elicit immediate behavior change (Bromberg-Martin, Matsumoto & Hikosaka, 2010; Gagné & Deci, 2005) leads to strongest effects. In contrast, neutral prosody elicited the weakest P2 amplitudes, that is messages and/or prosody that does not automatically trigger response for action might not be considered as relevant to the listener as messages that signal a requirement to action or participation through the tone of voice used. Notably, although controlling tones communicating an immediate need to respond elicited the strongest P2 amplitude, autonomy-supportive tones also elicited stronger responses as compared to neutral tones of voice, indicating support offered meaningful motivational content, but suggesting a softer quality to messages that motivate listeners through inviting opportunity through choice and encouragement when control is the comparison. Overall, these data nicely complement the attitudinal and emotional prosody literature which has repeatedly shown that emotional and attitudinal signals as conveyed through prosody are processed rapidly (e.g., Paulmann and Kotz, 2008b; Paulmann et al, 2013; Schirmer et al, 2013; Jiang and Pell, 2015). Crucially, it has been proposed that prosodic stimuli that are intrinsically relevant (e.g., signaling that immediate action is needed) may lead to enhanced processing efforts. This early evaluation has been argued to be supported by subcortical structures such as the amygdala or ventral striatum (c.f. Schirmer & Kotz, 2006). These structures are also implicated in the "salience network" (e.g., Menun & Uddin, 2010) linked to motivation processing (DiDomenico & Ryan, 2017). In short, it can be speculated that affective and motivational prosody might follow a similar on-line processing time-course as the two functions of prosody are also mediated by similar, overlapping brain networks.

Continuous monitoring of motivational cues

In addition to the early differentiation of motivational qualities and neutral, we also found an effect in the late potential window. Specifically, the different qualities of prosody were still processed differently at this point in time; however, there was an indication that the distribution of effects varied depending on motivational quality. While differences between controlling and neutral prosody were found at all electrode-sites, ERP differences in response to autonomy-supportive and neutral prosody were most pronounced at left (and to a lesser extent right) frontal electrode-sites. Finally, autonomy-supportive and controlling prosody led to differently modulated late potential component amplitudes at left (and to a lesser extent right) parietal electrode-sites. These findings suggest that after an initial evaluation of prosodic cues (P2 component), the extracted cues receive further attention to monitor the unfolding of motivational information. In other words, once motivational messages are detected and considered as relevant, they are evaluated more deeply as compared to neutral messages, as shown in larger late potential amplitudes for motivational as opposed to neutral prosodies. This observation fits well with previous research that links prosody processing processes in this later time-window to more enhanced, or attention demanding, cognitive processing (e.g., Kotz & Paulmann, 2011; Paulmann & Kotz, In Press; Schirmer & Kotz, 2006). Moreover, in addition to observing processing differences between neutral and motivational prosodies, we also observed differences between the two different psychological processes. We suggests that the different processing of the two motivational qualities in this later window lends support to the idea that the complexity of the two motivational qualities in terms of acoustic cues and their variability (Weinstein et al., 2018) requires distinct and elaborate evaluations at this point (see Zougkou et al., 2017). In short, the data nicely add to the growing body of evidence that listeners continuously monitor incoming speech not only to extract semantic information, but also to gauge the social intention of the speaker (c.f. neuro-cognitive models of prosody processing). Specifically, these later effects support the idea that in addition to early relevance or salience detection, listeners also engage in a “second pass analysis” of prosodic characteristics.

Interestingly, the differences in distribution between autonomy-supportive and controlling prosody also once more suggest that the two motivational qualities are processed, at least partly, by different neural mechanisms. It always needs to be highlighted that the spatial resolution of ERPs is low, but differences in distribution of effects observed here are still informative and suggest a partially different neural network is at play (Otten & Rugg, 2005) when processing autonomy-supportive and controlling prosody. Previous work had hypothesized that control and autonomy support are modulated by different brain networks based on P2 distributions (Zougkou et al., 2017), and specific brain circuits (Lee and Reeve, 2013). Here, we identify further evidence to support this hypothesis with data showing regional ERP hot spots when listening to autonomy-supportive as opposed to controlling (or neutral) prosody.

Prosody change detection

While the direct comparison of motivational and neutral prosody was informative, it was potentially confounded by the possibility that some information was temporarily prioritized during processing. This is particularly important if the temporal availability of motivation-relevant signaling auditory cues might vary between autonomy-support and control (Zougkou et al., 2017). In other words, stronger or more nuanced effects for one condition over another might be due to the point in time at which the listener received the motivational information. For instance, acoustic cues signaling autonomy support might be particularly subtle at sentence onset, while cues signaling control might be particularly strong. To account for this, the present study employed cross-splicing, which has been argued to allow for a more direct alignment of information processing (e.g., Kotz & Paulmann, 2007). Specifically, cross-spliced stimuli provide listeners with enough information to build up expectancy about how a sentence will continue in terms of relevant (in this case, motivational) information before new information (cross-spliced) that *mismatches* this expectation is introduced. In fact, it has been argued repeatedly that vocal communications are particularly prone to prediction mechanisms (see Bendixen, SanMiguel, & Schröger, 2012 for a review). This allows listeners to achieve appropriate goal-directed behavior (e.g., here, to carry out an action) and to allow for high flexibility during processing. In other words, listeners engage in predictive or anticipatory strategies to keep processing costs under control and to avoid possible (cognitive) overload. Here, we applied the cross-splicing paradigm to tap into these predictive mechanisms and to investigate motivation-specific processing.

Sense the change! Transitions from neutral to motivational prosody

The first splicing condition closely followed previous approaches from the emotional prosody literature (e.g., Kotz & Paulmann, 2007; Paulmann et al., 2012), and created prosodic contour deviations by splicing a prosodically neutral beginning to an autonomy-supportive or controlling sounding end. Both types of deviations elicited a broadly distributed positive ERP response in listeners shortly after the splicing-point (~400 ms), an effect which closely resembles the well-described PEP component reviewed earlier suggesting that transitions from neutral to motivational prosody are quickly detected irrespective of which specific motivational quality is conveyed. In the case of both motivations, listeners picked up on the change in speakers' tone of voice and engaged in more comprehensive re-analysis of cues to assess what this change might mean for them. Thus, these data confirm the simple prosodic mismatch effect that has been observed in both the emotional and attitudinal literature previously, including the observation that changes from neutral prosody are quickly detected and attended to, but the quality of prosody changed does not seem to affect this general detection mechanism (Paulmann & Kotz, 2008a). The only differences observed in previous studies seem to relate to latency and distribution; for example, transitions to emotional prosody are detected earlier than transitions to a different linguistic function (c.f. Paulmann et al., 2012) and effects are differently distributed.

Sense the tone: changing motivational styles while still talking

In a second condition, we specifically focused on the two motivational qualities of interest: autonomy support and control. We explored whether motivations are processed differently by splicing an autonomy-supportive beginning to a controlling ending and vice versa. If the two motivational qualities are processed differently (and relying on a partially different neural network as hypothesized earlier), we can expect ERP effects that clearly differ in terms of latency and distribution (Paulmann et al., 2012) and possibly even polarity (Kotz & Paulmann, 2007). Similar to transitions from neutral to motivational prosody described above, the second condition elicited a positive ERP component between 400-700ms after splicing point irrespective of motivational quality. Again, this effect nicely adds to the growing body of research that the PEP is a brain signature reflecting the detection of significant prosody changes possibly triggering re-analysis of acoustic cue patterns. Of more interest here is that transitions from autonomy-supportive to controlling prosody lead to an *additional* ERP effect in listeners, namely an early (~150 ms after splicing point), broadly distributed, short-lived negativity. Previous work has argued that a change in content occurring simultaneously to a change in prosody leads to quicker and faster detection of deviances and also engages additional neural structures (Paulmann & Kotz, 2008a). While semantic content was kept constant across conditions by presenting non-motivationally biasing sentences, it could be argued that the task which put the spotlight on prosody helped create a strong enough *context* expectation. Thus, past interpretations of the negativity could be extended not only to respondents registering and analyzing expected content violations, but also to their reactions to *context* violations. In fact, the probe detection task applied by Paulmann & Kotz (2008a) asked listeners to more strongly focus on content, possibly also helping listeners to build up expectations about content and quickly realizing when these were not fulfilled. Thus, it seems reasonable to assume that similar mechanisms were at play here when task focus helped listeners to pay attention to prosodic cues creating expectations about context development. The early onset of the negativity seems to indicate that listeners can quickly detect a clear deviation from expected content and/or context. The further observation that only transitions from autonomy-supportive to controlling prosody but not vice versa lead to such an early effect suggested that the change in the significance of the stimulus was predominantly driving this effect. That is, if the change required listeners' immediate and strong reaction, as controlling motivational communications tend to do, resources were quickly allocated to attend to the situation. This speculation receives support from previous reports that fail to identify effects for transitions from angry to neutral prosody (Chen et al., 2011), possibly because a change to a neutral sounding voice does not require an immediate response from the listener as neutral is considered to be a state of relaxation when compared to anger. The findings then also substantiate our previous reports of the relevance of controlling prosody as behavior of immediate action (Zougkou et al., 2017).

Conclusion

The present study aimed to further investigate the time-course underlying motivational prosody processing. Overall, observed findings confirm the hypothesis that different motivational qualities are assessed rapidly during sentence comprehension. The data nicely demonstrate that both autonomy-supportive and controlling prosody can be considered as "motivationally relevant" from an early time in processing at least when listeners pay particular attention to the tone of voice used. Later, this tagging process of motivationally relevant information is followed by more attention-demanding, in-depth analysis of motivational messages. The processing mechanisms identified for motivational prosody are similar to those linked to emotional prosody processing (e.g., Schirmer & Kotz, 2006; Kotz & Paulmann, 2011; Frühholz, Trost, Kotz, 2016) but latency and distribution of effects also further support the idea that emotions and motivations are conceptually and operationally distinct constructs expressed through overlapping but distinct prosodic cues (Weinstein et al., 2018; Zougkou et al., 2017). Finally, putting the spotlight on prosodic processing showed that messages from both motivational qualities are processed differently from non-motivational messages; however, directly comparing motivations in a cross-splicing design also confirmed the previously observed "preferential" processing for controlling information. Thus, the unique imprint of controlling prosody makes it nearly impossible for listeners to ignore it.

References

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

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636.

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: a review. *International Journal of Psychophysiology*, *83*(2), 120-131.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.0.14, 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), 815–34.

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

Chen, X., Zhao, L., Jiang, A., & Yang, Y. (2011). Event-related potential correlates of the expectancy violation effect during emotional prosody processing. *Biological Psychology*, *86*(3), 158-167.

Deci, E.L., Ryan, R.M. (1985). The general causality orientations scale: self-determination in personality. *Journal of Research in Personality, 19*(2), 109–34.

De Muynck, G. J., Vansteenkiste, M., Delrue, J., Aelterman, N., Haerens, L., & Soenens, B. (2017). The Effects of Feedback Valence and Style on Need Satisfaction, Self-Talk, and Perseverance among Tennis Players: An Experimental Study. *Journal of Sport and Exercise Psychology*, 1-38.

Di Domenico, S. I., & Ryan, R. M. (2017). The emerging neuroscience of intrinsic motivation: a new frontier in self-determination research. *Frontiers in human neuroscience*, *11*, 145.

Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research, 25*, 130–143.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport*, *12*(15), 3189-3191.

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.

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), 665–74.

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

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), 108–25.

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 Cognive and Affective Neuroscience, 8*, 538–545. doi: 10.1093/scan/nss029

Lee, W., & Reeve, J. (2017). Identifying the neural substrates of intrinsic motivation during task performance. *Cognitive, Affective, & Behavioral Neuroscience*, *17*(5), 939-953.

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), 1570–84.

Luyckx, K., Soenens, B., Berzonsky, M. D., Smits, I., Goossens, L., & Vansteenkiste, M. (2007). Information-oriented identity processing, identity consolidation, and well-being: The moderating role of autonomy, self-reflection, and self-rumination. *Personality and Individual Differences*, *43*(5), 1099-1111.

Mitchell, R.L., Ross, E.D. (2013). Attitudinal prosody: what we know and directions for future study. *Neuroscience & Biobehavioral Reviews, 37*(3), 471–9.

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

Otten, L.J., Rugg, M.D. (2005). *Interpreting event-related brain potentials*. In: Handy Todd, C., editors. *Event-Related Potentials: A Methods Handbook*. Cambridge: MIT Press.

Paulmann, S., & Kotz, S. A. (2008a). An ERP investigation on the temporal dynamics of emotional prosody and emotional semantics in pseudo- and lexical sentence context. *Brain and Language, 105*, 59–69.

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

Paulmann, S. (2015). *The neurocognition of prosody*. In: Hickok, G., Small, S., editors. *Neurobiology of Language*. San Diego: Elsevier.

Paulmann, S., Kotz, S.A. (In Press). *The Electrophysiology and Time-Course of*

Processing Vocal Emotion Expressions. In Belin, P., Frühholz, S., editors. *Oxford Handbook of Voice Perception*. Oxford: Oxford University Press.

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., 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., 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*, 1609–20.

Paulmann, S., Pell, M. D., & Kotz, S. A. (2008). Functional contributions of the basal ganglia to emotional prosody: Evidence from ERPs. *Brain Research*, *1217*, 171-178.

Paulmann, S., Vrijders, B., Weinstein, N., & Vansteenkiste, M. (2018). How parents motivate their children through prosody. In *Proc. 9th International Conference on Speech Prosody 2018* (pp. 250-254).

Picton, T. W., Goodman, W. S., & Bryce, D. P. (1970). Amplitude of evoked responses to tones of high intensity. *Acta oto-laryngologica*, *70*(2), 77-82.

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*, 657–74.

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.

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.

Schirmer, A., & Kotz, S. A. (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.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., & Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport*, *16*(16), 1781-1785.

Spreckelmeyer, K. N., Kutas, M., Urbach, T., Altenmüller, E., & Münte, T. F. (2009). Neural processing of vocal emotion and identity. *Brain and cognition*, *69*(1), 121-126.

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), 351–64.

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. In: Proceedings of the 7th Conference of 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 emotional and attitudinal prosody. *PLoS One, 10*(7), e0132947.

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.

Legend for Figure 1: The illustration shows event-related brain potentials in response to motivational stimuli at selected electrode-sites from 200 ms before up to 800 ms after stimulus onset/splicing point onset. Panel A displays P200 and late component effects in response to motivational prosody (e.g. "Come out for a walk with me" was spoken in three different prosodies). ERPs are time-locked to sentence onset. Panel B displays prosodic expectancy positivity (PEP) effects elicited for stimuli transitioning from neutral to motivational prosody (e.g., Come out [spoken in neutral] for a walk with me [spoken in controlling or supportive ways]. Panels C and D show effects in response to stimuli transitioning from controlling to autonomy-supportive (C) and autonomy-supportive to controlling prosody (D). ERPs displayed in panels (B), (C), and (D) are all time-locked to the onset of the splicing point (e.g., "for a walk…"). Negativity is plotted upwards. ERPs were filtered off-line with a 7-Hz low-pass filter for graphical display only (all statistical analyses were computed on non-filtered data).

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