Telling self-defining memories: An acoustic study of natural emotional speech productions

Véronique Delvaux[1], Audrey Lavallée[2,3], Fanny Degouis[3], Xavier Saloppé[2,4,5], Thierry Pham[3,4] & Jean-Louis Nandrino[2]

[1]UMONS, Language Research Institute & FNRS, Belgium
[2]University of Lille, SCALab, UMR CNRS 9193, France
[3]UMONS, Department of Legal Psychology, Mons, Belgium
[4]Research Center in Social Defense, Tournai, Belgium
[5]Psychiatric Hospital, Saint-Amand-les-Eaux, France

veronique.delvaux@umons.ac.be

Abstract

Vocal cues in emotion encoding are rarely studied based on real-life, naturalistic emotional speech. In the present study, 20 speakers (10 male, 10 female) aged 25 to 35 were recorded while orally telling 5 successive self-defining autobiographic memories (SDM). By definition, this task is highly emotional, although emotional load and emotion regulation are expected to vary across SDM. Seven acoustic parameters were extracted: MeanF0, MedianFo, StandardDeviationF0, MinF0, MaxF0, Duration and SpeechRate. All SDM were manually transcribed, then their emotional lexicon was analysed using Emotaix.

First, speech productions were examined in reference with SDM characteristics (specificity, integrative meaning and affective valence) as determined by 3 independent investigators. Results showed that overall the speech parameters did not change over the time course of the experiment, or as a function of integrative meaning. Specific memories were recounted at a higher speech rate and at greater length than non specific ones. SDM with positive affective valence were shorter and included less variability in fundamental frequency than negative SDM.

Second, emotionally-charged (positive vs. negative; high vs. low arousal) vs. emotionally-neutral utterances as to Emotaix classification were compared over all SDM. Only a few significant effects were observed, which led us to discuss the role of emotion regulation in the SDM task.

Index Terms: emotions, self-defining memories, emotion regulation, emotion encoding, autobiographical memory, speech production

1. Introduction

Although the literature on the encoding of emotions in speech has dramatically increased in recent years [e.g. 1,2,3,4,5], studies based on real-life, naturalistic emotional speech are still rare. This is due to a variety of reasons including ethical, practical and theoretical reasons. Indeed, it is ethically precarious to submit participants to experimental designs which purpose is to repeatedly elicit real, strong emotions of a large variety. Alternatively, one can rely on real-life speech utterances produced in selected contexts known to elicit emotional experiences. However, the vocal markers of emotions in such productions are typically hard to detect because the emotional effects are weak, vary greatly over individuals, and may be hindered by expression control due to display rules and strategic concern [5]. Moreover, the a posteriori classification of individual speech samples as representing the expression of a given emotion is highly dependent on the theoretical model of emotions underlying the classification.

In the present study, the participants were recorded while orally telling several self-defining memories (SDM). SDM are key components of autobiographical memory. They refer to vivid, emotionally intense and repeatedly recalled memories that concern lasting issues or unresolved conflicts. They help to maintain self-consistency and self-coherence, and also have a social dimension since people frequently describe themselves and their life stories by sharing SDM. SDM are classified based on several dimensions including specificity (the level of details of the recalled event), integration of meaning (the ability to update self-concept and personal goals by integrating experience from the recalled event) [6,7], and affective valence (positive, negative, neutral or mixed) [8]. By definition, recounting SDM is a highly emotional task, although emotional load and emotion regulation are expected to vary across SDM, potentially in a speaker-specific way. It thus constitutes both a challenge and an opportunity for the study of naturalistic emotional speech.

In the Memantemo research project (to which this study belongs), a multiparametric approach was adopted to study emotional processes in individuals with and without antisocial personality disorders. Face expressions, speech productions, as well as electrodermal activity and heart rate variability were simultaneously recorded while participants recalled SDM. Each SDM was characterized a posteriori by both expert clinicians and the participants themselves using a variety of tools. Extensive biographical, medical and psychological informations were also collected on the participants. The combination of several types of data allows for better control over some of the factors found in earlier studies to obscure vocal markers of emotions. For example, electrodermal activity and heart rate variability inform on the activity of the sympathetic and parasympathetic systems, which can be interpreted in terms of regulatory emotional processes [9,10]. In the context of the present study, emotion type was specified by a content analysis of the emotional lexicon used in each SDM.

This paper reports on a first exploratory study of the speech productions collected within the Memantemo project. The analysis is two-fold. First, speech productions were examined in reference with SDM characteristics (specificity, integration of meaning and affective valence) in order to investigate how vocal cues may reflect some properties of the emotional processes involved in autobiographic memory. Second, a content analysis of all the SDM was conducted using the Emotaix scenario so as to compare the acoustic parameters of emotionally-charged (positive vs. negative; high vs. low arousal) vs. emotionally-neutral utterances.

2. Material and methods

2.1. Task and participants

Twenty speakers (10 male, 10 female) aged 25 to 35 participated in the experiment. Of a higher education level, they were recruited through an advertisement on a social network. The participants showed no sign of having any serious psychological disorder or somatic illness, as per an interview with a psychologist.

Participants were asked to orally recount five SDM while their speech productions, facial expressions and psycho-physiological parameters were recorded. SDM were elicited following specific instructions: "You are invited to recall five events in your life. These events must be important in defining who you are. In other words, these memories should refer to events that help you to understand who you are as an individual. These events should also be events that you would share with someone if you wanted that person to understand you in a fundamental way. The events may be positive or negative memories; the only important aspect is that they should lead to strong feelings. The memories should be events that you have thought about many times. They should also be familiar to you like a picture you have looked at a lot, or a song you have learned by heart". Since one participant failed to recall a fifth SDM, the total of collected SDM was 99.

2.2. Coding of SDM

SDM were coded a posteriori (based on the video recordings) and independently by three investigators using Singer and Blagov's classification system and scoring manual for self-defining autobiographical memories [6]. They were coded as "specific" if they were a memory of a single specific brief event including details, and as "nonspecific" when they referred to a memory of long or repeated events. SDM were considered as "integrative" if memories were directly associated with learning about oneself, others or the environment, and "nonintegrative" if they were pure narratives without integration of the recalled experiences. Finally, SDM affective valence was coded as positive, negative, neutral or mixed (i.e. both negative and positive) based on the emotional vocabulary used by the participant [8]. After some training, inter-rater agreement proved substantial, as evidenced for by Cohen's kappa (specificity: k= .79; affective valence: k= .87; integrative meaning: k= .8).

2.3. Content analysis

All 99 SDM were manually transcribed, for a total of 5 hours 39 minutes of speech and 58956 words. A content analysis was performed using the EMOTAIX-Tropes text analysis software [11], then manually verified by two experts. The software automatically detects and counts emotional items, then classifies them based on their valence. More specifically, EMOTAIX comprises 2 x 28 basic emotional categories organized into three hierarchical levels on either side of a hedonic axis (positive and negative valence).

The present study is based on the 18 "super-categories" from the EMOTAIX classification, i.e. 9 of negative valence and 9 of positive valence (Table 1). Another, customized regroupment of these 18 categories was made in terms of arousal, i.e. high arousal vs. low arousal. Specifically, all 56 basic emotional categories were assigned an arousal value from 1 (low activation) to 9 (high activation), following norms established by [12], and the mean score over all relevant descriptors was used to classify each of the 18 super-categories into two groups: High arousal (mean score > 5) vs. Low arousal (mean score < 5). For example, kindness {goodness, softness, patience, humility} achieved a mean score of 3.96 and was thus classified into the "low arousal" group, whereas happiness {bliss, joy, laughter} achieved a mean score of 6.5, and was thus considered as a "high arousal" emotion.

Table 1: *Classification of the 18 emotional super-categories identified by Emotaix into two groups based on valence (positive vs. negative) or arousal (high vs.low)*

Valence-based classification (Emotaix)		**Arousal-based classification (Customized)**	
Positive	affection	High	affection
	kindness		happiness
	happiness		courage
	lucidity		hate
	spirit		agressiveness
	relief		madness
	satisfaction		depression
	courage		frustration
	calm		tension
Negative	hate	Low	kindness
	agressiveness		lucidity
	suffering		spirit
	madness		relief
	depression		satisfaction
	disorder		calm
	frustration		suffering
	fear		disorder
	tension		fear

2.4. Speech analysis

Orthographic transcriptions were aligned with speech productions using EasyAlign [13]. Acoustic analyses were carried out with Praat [14]. After basic signal filtering, fundamental frequency was extracted every 5 ms using gender-specific parameters. Seven acoustic parameters were computed: MeanF0, MedianFo, StandardDeviationF0 (below: SDF0), MinF0, MaxF0, Duration and SpeechRate (in words per second). These parameters were computed (i) over the entire SDM; (ii) separately for each utterance in which an emotional item was identified by EMOTAIX, as well as for the emotionally-neutral utterances forming the rest of the SDM. Statistical analyses were performed using SPSS© 20.

3. Results

3.1. Preliminary analysis

A repeated-measures multivariate analysis of variance was carried out with MeanF0, MedianFo, SDF0, MinF0, MaxF0, Duration and SpeechRate (computed over the entire SDM) as dependent variables, Gender (Male vs. Female) as between-subject factor and SDM Order (from 1 to 5) as within-subject factor. It revealed that neither SDMOrder nor its interaction with Gender had any statistical effect on the acoustic parameters. In contrast, Gender yielded significant differences in all F0-related parameters, i.e. MeanF0, MedianFo, SDF0, MinF0 and MaxF0. This was to be expected given the gender-related difference in F0 baseline: average MeanF0 was 104Hz for male speakers and 193Hz for female speakers. Male and female speakers did not significantly differ in terms of SDM duration or speech rate.

3.2. Speech productions as a function of SDM characteristics

Table 2 presents descriptive statistics on the seven acoustic parameters computed over the entire SDM (N=99), as a function of the SDM classification in terms of Integration, Specificity and Valence. Nonparametric Mann-Whitney tests were carried out in order to test for significant differences between integrative vs. nonintegrative, specific vs. nonspecific and positive vs. negative SDM.

Concerning valence, positive SDM proved significantly shorter than negative SDM (139s vs. 252s; U= 156, p< .05). Two other dependent variables achieved p-levels close to .05: SDF0 and MaxF0. MaxF0 was lower for positive SDM than for negative SDM (242Hz vs. 279Hz; U= 169, p= .051). Similarly, SDF0 was lower for positive SDM than for negative SDM (17Hz vs. 22Hz; U= 170; p= .054). Regarding specificity, speech rate was found significantly higher for specific (3 words/sec) vs. nonspecific SDM (2.7 words/sec): U= 754, p< .05). Specific SDM were also almost twice as long as nonspecific ones (240s vs. 130s): U= 548, p< .001. As to integration of meaning, there were no significant variations in the seven acoustic parameters as a function of this factor.

3.3. The expression of emotions in SDM

The following analyses are based on acoustic parameters which were computed over portions of SDM, as per the EMOTAIX analysis, i.e. either the "emotionally-neutral" utterances or the "emotionally-charged" utterances (each assigned to one out of 18 emotional categories, cf. Table 1). Data were aggregated for each category over the 5 SDM of each participant (N=20).

First, we compared emotional portions of the SDM (all 18 emotional categories) with nonemotional ones ("EMO" vs. "nonEMO" below). Descriptive statistics are given in Table 3. Wilcoxon signed-ranks tests were carried out in order to test for significant differences between related samples (EMO and nonEMO) in terms of MinF0, MaxFO, MeanF0, SDF0, MedF0, Duration and SpeechRate. They revealed that EMO utterances were significantly shorter (Z= -3.92, p< .001) and produced at a higher speech rate (Z= -3.81, p< .001) than nonEMO samples. In terms of F0 frequency, there was no difference between EMO and nonEMO in terms of SDF0 or MeanF0, and a significant difference of negligible size in terms of MedianF0 (147Hz vs. 145Hz, Z= -2.91, p< .05). Besides, MinF0 was significantly higher in EMO than in nonEMO (121Hz vs. 107Hz, Z= -3.92, p< .05), while MaxF0 was significantly lower in EMO than in nonEMO (205Hz vs. 257Hz, Z= -3.92, p< .001).

Second, we focused on emotionally-charged utterances only, comparing between utterances containing a lexical item of positive valence (9 emotional categories) and those containing a lexical item of negative valence (9 emotional categories) according to the EMOTAIX scenario (Table 1, valence-based classification). We thus grouped data from the 18 emotional categories identified by EMOTAIX onto two groups based on valence: positive vs. negative ("POS" vs. "NEG" below). According to Wilcoxon signed-ranks tests for related samples, only MeanF0 and MedianF0 significantly varied between POS and NEG but the size of the effects was negligible (less than 1,5Hz in both cases).

Third, we grouped data from the 18 emotional categories identified by EMOTAIX onto two groups based on arousal: high vs. low ("HIGH" vs. "LOW" below). Wilcoxon signed-ranks tests were carried out in order to test for significant differences between related samples, but they revealed no difference between HIGH and LOW in any of the acoustic parameters.

4. Discussion

In this study, we performed an acoustic study of speech productions recorded during a task evoking strong emotions associated with autobiographic memory, i.e. the retrieval of self-defining memories. Based on the relevant literature, [e.g. 1,2,3,4,5] emotional speech utterances, especially those involving high-arousal emotions, were expected to result in increased F0 baseline, F0 variability and speech rate.

First, speech productions were examined in reference with SDM characteristics (specificity, integrative meaning and affective valence) and order of recall. Results showed that overall the acoustic parameters did not change over the time course of the experiment, from SDM1 to SDM5. Specific memories were recounted at a higher speech rate and at greater length than non specific (general) ones. Higher speech rate is consistent with specific memories containing more emotional details. Concerning affective valence, positive SDM were shorter and included less variability in fundamental frequency than negative SDM. Integration of meaning did not yield any significant difference in the acoustic parameters. This latter result was unexpected since in a recent related work nonintegrative SDM were shown to elicit more skin conductance [15], which is typically associated with increased sympathetic activity reflecting high arousal.

Secondly, the speech productions from the 5 SDM of each speaker were considered together, and their acoustic parameters were analyzed with regards to their emotional content as specified by EMOTAIX. Several differences proved statistically significant between emotionally-charged and emotionally-neutral utterances. However, some of them were not really meaningful. For example, the observation that EMO utterances had shorter duration and higher speech rate might be attributed to the fact that emotional portions of the SDM were targeted as single utterances, whereas the nonemotional portions were typically made of several consecutive utterances interspersed with short pauses. Moreover, the grouping of emotionally-charged utterances as to either valence (POS vs. NEG) or arousal (HIGH vs. LOW) did not allow for significant acoustic differences to emerge.

Altogether, the emotional effects measured in the present study might be considered of limited impact, or in some instances going against predictions. Several reasons may be offered as an explanation, some regarding the task itself, others regarding the analyses performed here.

Table 2: *Descriptive statistics. MinF0, MaxF0, MeanF0, MedianFo, SDF0, Duration and SpeechRate (computed over the entire SDM) as a function of SDM characteristics: Integration, Specificity, Valence.*

			MinF0	MaxF0	MeanF0	sdF0	MedianF0	Duration	SpeechRate
Integration	Integrative	Mean	105	259,4	148,8	19,7	144,5	224,6	2,9
	(N=53)	SD	30,4	79,4	47,8	9,1	46,4	152,7	0,5
	Nonintegrative	Mean	107,6	255,1	148,1	18,6	143,8	183,4	2,9
	(N=46)	SD	31	77,9	44,9	6,8	43,8	153,8	0,5
Specificity	Specific	Mean	103,6	253,8	144,6	18,8	140,5	239,7	3
	(N=68)	SD	30,4	80,8	44,9	8,2	43,6	168,3	0,5
	Nonspecific	Mean	111,9	265,4	156,9	20,1	152	130,2	2,7
	(N=31)	SD	30,7	73,3	48,7	8	47,6	74,6	0,5
Valence	Positive	Mean	103,4	241,8	144,1	17	140,2	138,6	3
	(N=19)	SD	31,6	71,6	46,1	6	45	54,5	0,4
	Negative	Mean	112,9	278,6	159,6	21,5	154,5	252,2	2,9
	(N=27)	SD	30,6	75,6	44,9	8,7	43,4	203	0,4
	Neutral	Mean	114,2	264,9	159,4	18,8	154,9	159,2	2,8
	(N=12)	SD	32,7	78,4	47,7	5,7	46,7	154,5	0,5
	Mixed	Mean	100,8	248,5	139,9	18,8	136	219,1	2,9
	(N=41)	SD	29,2	82,5	46,2	9	45,1	137,4	0,6

Table 3: *Descriptive statistics. MinF0, MaxF0, MeanF0, MedianFo, SDF0, Duration and SpeechRate as a function of emotional content over all SDM.*

			MinF0	MaxF0	MeanF0	sdF0	MedianF0	Duration	SpeechRate
Emotional charge	Emotional	Mean	121	204,6	150,1	17,6	147,2	4,2	3,5
	N=20	SD	37,7	58,7	47,3	5,1	47,1	1	0,4
	Nonemotional	Mean	107,5	257	149,7	19,7	145,1	156	2,7
	N=20	SD	31,3	78,4	47,8	7,9	46,3	88,1	0,5

Concerning the task, although retrieving SDM is typically described as a highly emotional task [e.g. 16,17], it might result in emotional effects that are too weak to be detected in speech, because they are not associated with direct emotional experiences but only with the emotional content of events retrieved from memory. Moreover, as a rule real-life, naturalistic speech tokens do not give direct access to emotions because their expression is controlled in accordance with display rules (e.g. politeness) and strategic concerns [5]. In fact, the originality of the experimental task is that emotion expression in SDM can be attributed to the retrieval of the original emotional experience and/or to the emotional *regulation* of the retrieved memories [18]. For example, emotional regulation may be responsible for one result that could appear as counter-intuitive on first examination, i.e. the fact that EMO utterances had higher MinF0 and lower MaxF0, thus decreased F0 range, in comparison with nonEMO utterances. In sum, self-defining memories are particular memories, at the crossroad between cognitive and emotional processes.

Other reasons for the limited effects sometimes observed here may regard the way emotionally-charged utterances were detected then specified, i.e. via a content analysis supervised by EMOTAIX. A first concern relates to the fact that the EMOTAIX scenario is basically organized across a hedonic axis (positive vs. negative valence), while arousal might be the primary dimension to correlate with F0 and speech rate variations. Note however that we found no difference between high-arousal and low-arousal emotional utterances in the extracted acoustic parameters. More fundamentally, it is possible that EMOTAIX appropriately detect the utterances which *could* carry emotional effects, without them being actually present every time (due to some of the factors detailed above). In this line of thought, we plan to reverse the perspective on our data: future analyses will entail a "bottom-up strategy", starting from the distributional properties of the acoustic parameters themselves in order to identify potentially emotional utterances in each SDM, then checking whether or not these utterances were indeed more frequently identified by EMOTAIX as emotional speech.

5. Acknowledgements

Véronique Delvaux is a research associate of the Fund for Scientific Research – FNRS, Belgium.

The authors wish to thank the organisms supporting the Memantemo project: the French Agence Nationale de la Recherche (ANR-11-EQPX-0023), the FEDER SCV-IrDIVE European program, the Hauts-de-France Region and the University of Mons. Special thanks to Florian Sanssen for his valuable contribution to data collection.

6. References

[1] T., Johnstone, K.R. Scherer, "Vocal communication of emotion". In: Lewis, M., Haviland, J. (Eds.), *Handbook of emotion*, second ed., New York: Guilford, pp. 220– 235, 2000.

[2] K.R. Scherer, "Vocal communication of emotion: a review of research paradigms", *Speech communication*, vol. 40, pp. 227-256, 2003.

[3] T. Bänziger, K.R. Scherer, "The role of intonation in emotional expressions", *Speech Communication,* vol. 46, pp. 252-267, 2005.

[4] K. R. Scherer, "Component models of emotion can inform the quest for emotional competence", In G. Matthews, M. Zeidner, & R. D. Roberts (Eds), *The science of emotional intelligence*: *Knows and unknows,* New York: Oxford University Press, pp. 101-126, 2007.

[5] K. R. Scherer, "Vocal markers of emotion: Comparing induction and acting elicitation", *Computer Speech & Language*, vol. 27, no. 1, pp. 40–58 , 2013.

[6] J.A. Singer, P.S. Blagov, "Classification system and scoring manual for self-defining autobiographical memories". Unpubl Manuscr Conn Coll, 2000.

[7] P.S. Blagov, J.A. Singer, "Four dimensions of self-defining memories (specificity, meaning, content, and affect) and their relationships to self-restraint, distress, and repressive defensiveness", *J Pers*, vol. 72, pp. 481–511, 2004.

[8] J.L. Nandrino, M.C. Gandolphe, "Characterization of Self-Defining Memories in Individuals with Severe Alcohol Use Disorders After Mid-Term Abstinence: The Impact of the Emotional Valence of Memories", *Alcohol Clin Exp Res*, vol. 41, pp. 1484–1491, 2017.

[9] B.M. Appelhans, L.J. Luecken, "Heart rate variability as an index of regulated emotional responding", *Rev Gen Psychol*, vol. 10, 229, 2006.

[10] D. Tranel, "Electrodermal activity in cognitive neuroscience: neuroanatomical and neuropsychological correlates", *Cogn Neurosci Emot*, pp. 192–224, 2000.

[11] A. Piolat, R. Bannour, "An example of text analysis software (EMOTAIX-Tropes) use: The influence of anxiety on expressive writing", *Current Psychology Letters*, vol. 25, Issue 2, 2009.

[12] A.B. Warriner, V. Kuperman, M. Brysbaert, "Norms of valence, arousal, and dominance for 13,915 English lemmas", *Behavior research methods*, vol. 45, no. 4, pp. 1191-1207, 2013.

[13] J. Goldman, "EasyAlign: An automatic phonetic alignment tool under praat". *Interspeech'11, 12th annual conference of the international speech communication association*, 2011.

[14] P. Boersma, D. Weenink, D. "Praat: doing phonetics by computer" [Computer program]. Version 6.0.50, retrieved 31 March 2019 from http://www.praat.org/

[15] A. Lavallée, X. Saloppé, M.C. Gandolphe, L. Ott, T. Pham., J.L. Nandrino (in revision). "What effort for retrieving self-defining memories? Specific autonomic responses for integrative and non-integrative memories"

[16] A.R. Sutin, R.W. Robins, "Continuity and correlates of emotions and motives in self-defining memories", *J Pers*, vol. 73, pp. 793–824, 2005.

[17] W. Wood, M. Conway, "Subjective Impact, Meaning Making, and Current and Recalled Emotions for Self-Defining Memories", *J Pers*, vol. 74, pp. 811–46, 2006.

[18] M.C. Gandolphe, J.L. Nandrino, G. Delelis, C. Ducro, A. Lavallee, X. Saloppé, A.A. Moustafa, M. El Haj, "Positive facial expressions during retrieval of self-defining memories", *Journal of Integrative Neuroscience*, pp. 1–10, 2017.