Valence of Emotional Memories: A study of lexical and acoustic features in older adult affective speech production

Author:
Ellen Tournier (s1013645)

Supervisors:
Dr. K. P. Truong
D. S. Nazareth
Dr. E. Janse

June 20th, 2019

MA Taal- en Spraakpathologie
Radboud University of Nijmegen

“The most beautiful emotion we can experience is the mystical. It is the power of all true art and science. He to whom this emotion is a stranger, who can no longer wonder and stand rapt in awe, is as good as dead, and his eyes are dimmed.”

– Albert Einstein[1]

[1] Einstein, A. (1931). The world as I see it. *Forum and Century, 84*(13), 193-194.

Contents

Preface

Before you lies the master's thesis *Valence of Emotional Memories: A study of lexical and acoustic features in older adult affective speech production* which I wrote to fulfill the graduation requirements of the 'Taal- en Spraakpathologie' (TSP; meaning Language and Speech Pathology) program at the Radboud University Nijmegen (RU). I was engaged in researching and writing this thesis from February to June 2019. My thesis is written as part of a bigger project on vocal expressions of emotion in older adult speech, that is undertaken at the University of Twente (UT). During my internship at the UT, I learned valuable lessons about the practical side of conducting research that I had never thought of before, but which made science a lot more intriguing and alluring to me (which I never would have thought before, either).

First of all, I want to thank my first supervisor at the UT, Khiet P. Truong of the Human Media Interaction department. She encouraged me to develop a novel scheme for measuring emotional memories, the Valence of Emotional Memories scale (VEM). I could not have been this proud of this achievement without our weekly meetings and Praat workshops. I always felt free to ask questions and even when I came to her for a different working space, she was kind and understanding and showed me other places to sit and write my thesis. Secondly, I want to thank Deniece S. Nazareth of the Psychology, Health & Technology department at the UT. She led the original project on vocal emotional expressions of older adults and helped me out in different ways while transcribing and annotating audio data. The sentiment analysis would probably not have been conducted without her. I also owe her a big thanks for giving me the opportunity to contribute to her Interspeech 2019 paper, addressing the emotional valence in older adult affective speech. This chance really encouraged my enthusiasm for science and researching.

Furthermore, I want to thank my supervisor at the RU, Esther Janse, who was closely involved in the progress of my thesis, as well. I am very thankful for the multiple Skype meetings and fast email correspondences, especially when it came to the statistical analysis. Not only am I grateful to her for being a helping hand throughout my internship, but also for the guidance over the past two years during the TSP premaster's and master's program at the RU. During this program I met two friends, Sabine and Julia, who also managed to help me and keep me sane and motivated in times of stress. Overall, this master's degree and internship have brought me a lot to cherish and be proud of.

I hope you enjoy your reading.

Ellen Tournier

Deventer, June 20th, 2019

Abstract

Analyzing emotional valence in affective spontaneous speech production remains a challenging task. The current study aims at lexical and acoustic analyses of valence of older adult's autobiographical memories. A novel valence coding scheme is presented due to the subjective and personal nature of memories and is used to annotate the spontaneous affective speech of 12 older adults (M = 73.00, SD = 5.26). These findings were compared to the results of lexical and acoustic analyses of the same speech samples. A regression analysis showed that valence could successfully be predicted by lexical and acoustic features. A multivariate analysis of variance showed a significant relation between lexical features (word usage) and four acoustic features. In conclusion, this study provides a deeper understanding of valence in spontaneous affective speech production of older adults.

Key words: acoustic features, affective speech, emotions, lexical features, life events, memories, older adults, valence.

1. Introduction

Over the past few years, research on affective speech production started to shift its focus from simulated (e.g., Laukka, Juslin, & Bresin, 2005) to spontaneous speech production (e.g., Tahon, Degottex, & Devillers, 2012). This is a positive but complex development, because the elicitation of spontaneous speech must be robust to variables that are hard to control, such as the type of speaker or the recording conditions. From the affective speech studies that did elicit spontaneous speech, we know that emotional arousal (calm – agitated) has a compelling and consistent effect on vocal expression (Juslin & Scherer, 2005). Findings on emotional valence (positive – negative), on the other hand, are much more inconsistent and often contradictory (Juslin & Scherer, 2005; Scherer & Oshinsky, 1977). In addition, these results only originate from younger adult speech, while older adult speech should be taken into consideration as well, because it has been shown that emotions are not experienced the same way throughout life. When people grow older, the types of emotions that are experienced are generally of a higher valence (Pasupathi, Carstensen, Turk-Charles, & Tsai, 1998). To this extent, much more research has to be conducted to clarify the role of valence in spontaneous affective speech production of older adults.

It seems that measuring the valence of spontaneous affective speech production comes with challenges. Studies that developed ways to measure the valence of lexical speech features (word usage) of spontaneous speech (De Smedt & Daelemans, 2012b; Moors et al., 2013) show that it is not enough to look at just the (combinations of) emotional key words; the meaning of sentences, and therefore the valence of speech, gets lost in this type of analysis (De Smedt & Daelemans, 2012b). Other studies focused on the examination of the valence of acoustic speech features, such as pitch, intensity, and speech rate (e.g., Goudbeek & Scherer, 2010; Schröder, Cowie, Douglas-Cowie, Westerdijk, & Gielen, 2001; Tahon et al., 2012). The findings of these and additional studies on the valence of acoustic features are not homogeneous and do not constitute a solid foundation for future research. In the current study, spontaneous affective speech was elicited by asking older adults to share their emotional memories (i.e., sad and happy) using autobiographical memory recall. To measure valence in spontaneous affective speech, labels that indicate the degree of valence are needed for the annotation of speech samples. Existing annotating schemes (e.g., Aubergé, Audibert, & Rilliard, 2006; Kanluan, Grimm, & Kroschel, 2008), however, do not consider the complexity of emotional memories. Therefore, the current study proposes a novel scheme for annotating the valence of emotional memories in spontaneous affective speech.

The contribution of this study is three-fold. First of all, the field of spontaneous affective speech will be explored by focusing on the valence of various lexical and acoustic speech features, since the valence dimension is much less represented in emotional research than the arousal dimension. Secondly, the spontaneous speech production of older adults will be examined as a way to broaden the research field of affective speech. Lastly, a novel coding scheme will be presented to annotate the valence of emotional memories in spontaneous speech, that, in contrast to existing annotation schemes, accounts for the highly personal and subjective nature of memories. The main research question of this study is as follows: *How can valence be measured in spontaneous emotional speech production of healthy older adults (≥ 65 years of age)?*

The outline of the current study is as follows: chapter 2 focuses on previous research on emotions, autobiographical memories, and lexical and acoustic speech features. Section 2.1 examines the elicitation and measurement of emotions and the effect of aging on the experience of emotions. In section 2.2, emotions will be associated with autobiographical memories and their connection to life scripts and life events. The description of life events will form the basis for a novel developed valence scheme for annotating spontaneous emotional memories. Sections 2.3 and 2.4 will elaborate research on lexical and acoustic correlates of emotions, respectively. In section 2.5, the research questions and hypotheses will be described.

Chapter 3 focuses on the methods. The novel valence scheme, the Valence of Emotional Memories scale (VEM), will be introduced in section 3.2.2. Chapter 4 will show the results, after which they will be discussed in chapter 5. In this chapter, the limitations of the current study and suggestions for future research will be discussed as well. In the final chapter, chapter 6, the conclusion will be presented.

2. Background

In this chapter, an overview of previous research will be presented. First of all, the research field of emotions is explored by defining and classifying emotions and describing ways in which they can be elicited and measured. The effect of aging on experiencing and expressing emotions will briefly be discussed. Second, research on emotional autobiographical memories will be discussed, as well as how they are related to cultural life scripts, life events, and affective speech. Third, the connection will be made between lexical and acoustic speech features and various emotional dimensions, particularly emotional valence. Lastly, the research questions and hypotheses of the current study will be outlined.

2.1 Emotion

2.1.1 What are emotions?

Defining 'emotion' is a problem that science has been struggling with for a long time. Since James (1884) tried to answer the question "What is an emotion?" a debate has been started and continued with no end in sight. Currently, one of the most cited definitions of emotion (e.g., Barrett, 2016; Jacob-Dazarola, Ortíz-Nicolás, & Cárdenas, 2016; Shuman & Scherer, 2015) is that of the component process theory of Scherer (1982). In his view, emotion is defined as "an episode of interrelated, synchronized changes in the states of […] organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism" (Scherer, 2005, p. 697). It is important to distinguish emotions from related affective phenomena, like moods, attitudes, preferences, and affect dispositions. To make this distinction, Scherer (2005) proposed seven design features by which emotions can be discriminated from other affective phenomena. One of the most defining features of emotions is that they are *about* something: most of the time they are elicited by an event that is actually occurring, remembered, or imagined (e.g., succeeding at a task, remembering a big success, or living up to an upcoming chance to succeed). This is called the event focus of the emotion. However, people are not always consciously aware of the events that have stimulated the emotion (Shuman & Scherer, 2015). The event focus is linked to a second design feature: emotions are appraisal driven. This means that the particular event and its consequences must be relevant to someone for the emotion to be experienced. If we do not care about something, we do not generally get emotional about it. Furthermore, emotions can change very quickly, often because of new information or due to a rapidly changing event (e.g., an important meeting at work goes very well, until you misspeak and embarrass yourself). For that matter, emotions need to have a relatively short duration in order to be able to respond instantly to new situations. For example, typical anger and joy episodes last between 15 and 60 minutes, and fear episodes up to 15 minutes (Verduyn, Delvaux, Van Coillie, Tuerlinckx, & Van Mechelen, 2009). Other design features are response synchronization (all organismic subsystems must contribute to the response preparation to events), behavioral impact (emotions have a strong effect on human behavior), and intensity (emotions have a relatively high intensity).

Based on these design features, emotions differ from moods in the sense that moods are considered as diffuse affect states that often emerge without a clear cause. In other words, they do not have an event focus like emotions do. Moods are generally of low intensity, but may last for a longer period of time. Examples are being cheerful, gloomy, or depressed (Scherer, 2005). Attitudes can be described as beliefs and predispositions towards something, for example hate or desire. An attitude can have an event focus, but this can also be a thing, person, or group or category of individuals instead of an event or situation. Attitudes are not per definition driven by appraisals, although they can become more prominent when thinking of the attitude object. Like attitudes, preferences have an event focus, but they are rather stable and unspecific. They can be described as judgments in the sense of liking or disliking a stimulus in a relatively low intensity. Lastly, emotions can be distinguished from affect dispositions. These dispositions describe the tendency of a person to experience certain moods more frequently, like nervousness, anxiety, or jealousy. Affect dispositions also include emotional pathology; while it is quite normal to be in a depressed mood, being depressed all the time could hint to an affective disturbance (Scherer, 2005).

Not only is it essential to characterize various affective phenomena, including emotions, it is also important to make a distinction between different types of emotions: utilitarian and aesthetic emotions (Scherer, 2004). Utilitarian emotions, also called basic/discrete (Ekman, 1992; Izard, 2007) or modal (Scherer, 1994) emotions, are the types of emotions that are innate, universal, and can help us adapt to events that have essential consequences for our wellbeing, such as anger, sadness, and happiness. Aesthetic emotions are not primitive and are not required to immediately adapt to a certain situation. This type of emotion is produced by the appreciation of intrinsic qualities and can be experienced when looking at art or nature or listening to music (Ellsworth & Scherer, 2003). Examples of such emotions are admiration, bliss, and fascination (Scherer, 2005).

2.1.2 Classifying emotions

There are various emotion theories that try to describe, explain, and classify emotions. Two fundamental viewpoints are the discrete approach (e.g., Ekman, 1992; Izard, 2007) and the dimensional approach (e.g., Barrett, 2006b; Laukka et al., 2005; Russell, 2003) on emotions. The discrete approach describes emotions as categorically different phenomena from perceptions and cognitions, and each emotion differs from every other emotion. In this approach, certain emotions (anger, sadness, fear, disgust, surprise, and happiness) are assumed to be primitive and universal in humans, meaning that anyone can experience and recognize this emotion (Barrett, 2006b). However, there is no consensus over which emotions are in fact discrete and which are not (Ortony & Turner, 1990). Furthermore, it is assumed that each emotion has a 'physical fingerprint', which contains the details of that particular emotion, such as facial expression and a pattern of autonomic nervous system activity. However, research findings do not confirm a specific physical fingerprint for each emotion (Barrett, 2006a, 2006b), nor are the assumed universal facial movements supported by studies that use facial electromyography (see Barrett, 2006a, 2006b; Russell, Bachorowski, & Fernández-Dols, 2003). Moreover, discrete emotion perception studies are capable of showing robust and replicable findings, but only when using a forced-choice task in which participants are shown a posed face or body or listen to a caricatured vocalization, and then are provided with a small set of emotion words from which they choose the correct label. Findings generally suggest that participants choose the correct response more often than chance, leading to claims that emotions are universally recognized (Elfenbein & Ambady, 2002). But when the participant is free to label the emotion cue or the conceptual context is removed, agreement rates drop significantly (see Barrett, 2011a, 2011b). So far, the discrete approach and its assumptions have been largely unvalidated.

The dimensional approach describes emotions by their position in a multidimensional space formed by (at least) valence and arousal (Russell, 2003). Valence represents the attractiveness (positive valence) or aversiveness (negative valence) towards an event, object, or situation leading to the emotion. Arousal describes the degree of excitement (calm versus aroused) elicited by the object(s) of emotion (Frijda, 1986). Although the dimensions of valence and arousal have obtained the most attention in the study of emotions, there has been considerable disagreement about the number and nature of the dimensions that provide an optimal framework for describing emotions. For example, other proposed dimensions are potency/control, intensity, power, and unpredictability (Fontaine, Scherer, Roesch, & Ellsworth, 2007; Goudbeek & Scherer, 2010; Laukka et al., 2005, Schröder et al., 2001). These dimensions are proposed because two-dimensional models, such as the valence-arousal model (e.g., Yik, Russell, & Feldman-Barrett, 1999; see Figure 1), are not able to fulfill the need of a dimensional space that accounts for all similarities and differences in emotional experience (Fontaine et al., 2007). For example, anger can be characterized as a high-arousal, negative state, but so can fear, disgust, and a variety of other emotions (Barrett, 2016).

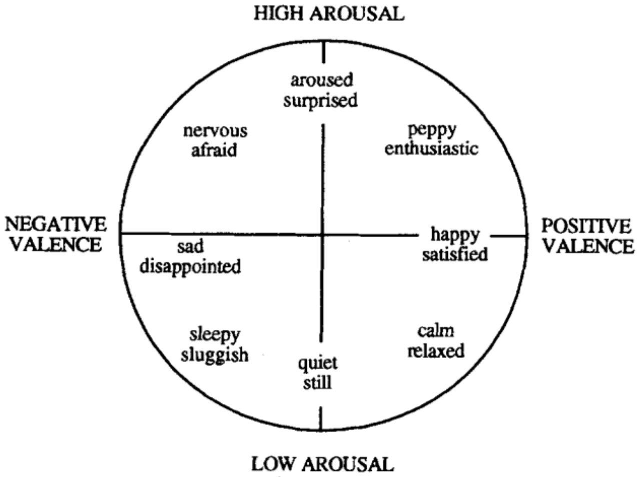


Figure 1. *Example of a valence-arousal model of emotions (taken from Feldman, 1993).*

2.1.3 Eliciting emotions

There are multiple ways to collect the experience of emotions for affective research. For example, eliciting *behavior* or a *feeling* of an emotion can be accomplished by presenting participants with films (e.g., Larsen, McGraw, & Cacioppo, 2001), music, (e.g., Hunter, Schellenberg, & Schimmack 2008), pictures (e.g., Schimmack & Colcombe, 2007), or advertisements (e.g., Andrade & Cohen, 2007). Emotions are also represented in emotional speech production. Most studies on vocal affect expression (e.g., Goudbeek & Scherer, 2010; Laukka et al., 2005) use recordings of emotion portrayals by professional actors. Usually, actors are asked to perform particular verbal material while portraying a set of discrete emotions, typically with high intensity (Van Bezooijen, 1984). There are advantages to this method, like experimental control, the production of strong voice effects, and the achievement of good sound quality on the recordings (Juslin & Scherer, 2005). The big disadvantage, however, is the question if actor portrayals are representative of spontaneous, natural affective speech. With the use of natural voice expressions there is an increased likelihood of obtaining ecologically valid speech samples and the preservation of the natural context of vocal expression (Juslin & Scherer, 2005). The most serious problem with obtaining natural affective speech is the difficulty of determining exactly which emotion is felt or portrayed by the speaker. Different persons react differently to the same situation, so if possible, studies should control for this factor. Furthermore, emotion detection in real-life conditions can be a big challenge, because it must be robust to variables that cannot be controlled, like the type of speaker (e.g., age and voice quality), the recording conditions (e.g., room acoustics and microphone quality), and the type of emotion that is elicited (Tahon et al., 2012).

Understandably, much fewer studies have worked with spontaneous speech, because it is much harder to obtain. One example of a study that has aimed to elicit spontaneous emotions in speech, is that of Tahon and colleagues (2012). They used the speech data of older people, that was evoked by interacting with a social robot that assists people at home in everyday activities. The data was then used for measuring the acoustic features of affective speech. Other uses of natural vocal expression involve recordings of conversations in psychotherapy (e.g., Roessler & Lester, 1976), with a focus on affective states like stress and depression, or recordings of speech samples of TV or radio (e.g., Douglas-Cowie, Campbell, Cowie, & Roach, 2003; Schröder et al., 2001).

2.1.4 Measuring emotions

When studies succeed in eliciting (the right) emotions, another challenge arises: the measurement of emotions. Objectifying and measuring emotions is difficult, especially in spontaneous settings, since emotions are highly personal and not every individual experiences or expresses the same emotion in a particular situation. Most affective studies (e.g., Fontaine et al., 2007) measure emotions according to one or more of the dimensions of the dimensional approach (Russell, 2003). There are various ways of measuring these emotion dimensions, all with their own advantages and disadvantages (see Mauss & Robinson, 2009). Self-reports, for example, are a popular way of measuring the valence and arousal of emotions in a subjective manner (Mauss & Robinson, 2009). Questionnaires are answered by individuals who experience different emotional states when presented with, for example, characters or pictures of faces that represent particular emotions. The degree to which self-reports are valid varies by the type of self-reports. Generally, self-reports of current emotional experiences seem fairly valid (Mauss & Robinson, 2009). Furthermore, facial movements and behavior seem to be valid measurements for emotion experience. In this case, facial electromyography is often used to determine the valence of the emotion (Barrett, 2006a, 2006b; Russell et al., 2003).

Another way to recognize and measure emotions is through acoustic speech features (Jacob-Dazarola et al., 2016). Studies on the vocal characteristics of emotional speech production often include acoustic components like fundamental frequency (F_0), intensity, and tempo (e.g., Laukka et al., 2005; Scherer, 1972; Scherer & Oshinsky, 1977; Schröder et al., 2001). It has been shown that arousal has a compelling effect on vocal expression (Banse & Scherer, 1996), often overpowering the effects of valence or other dimensions. One example that shows the distress in vocal recognition of arousal and valence comes from the study of Johnstone and Scherer (2000). They showed that, even though anger and joy are similar in arousal but different in valence, both emotions have been linked to comparable vocal pitch and vocal amplitude. Contradictory results like these show that there is a growing need for a better way to measure spontaneous emotions, particularly the *valence* of those emotions. In section 2.4, the acoustic features of affective speech production will be discussed in more detail.

2.1.5 Effect of age on emotion

Emotions are not experienced in the same way or in the same intensity throughout life. The early part of life is typically characterized by the acquisition of emotional specificity and efficiency in emotion regulation (Pasupathi et al., 1998). For example, in early life, infants simply appraise stimuli as something they like or dislike. By the age of four or five, children have not only acquired more differentiated emotional states such as anger or sadness, but they have also begun to acquire more complex emotions, like guilt or shame. From then on, however, research on the development of emotional functioning lessens. Most studies on emotional functioning have focused on young adults, leaving gaps in our knowledge about adolescent development of emotional functioning and the development beyond early adulthood. However, researchers have paid some attention to emotions in later life. When comparing the emotional data in younger and older adults, it seems that various aspects of emotional functioning are influenced by aging. As we grow older, emotional functioning appears to be largely spared from decline, and there is evidence that it may in fact improve. The types of emotions experienced by older adults are generally more positive and less negative (Pasupathi et al., 1998). For example, studies have suggested that anger may become less frequent or less intense in later life (e.g., Birditt & Fingerman, 2003, Schieman, 1999). This phenomenon is also known as the positivity effect (Kennedy, Mather, & Carstensen, 2004): an “age-related information processing bias towards positive versus negative information" (Reed, Chan, & Mikels, 2014, p. 1). According to the differential emotions theory (as cited in Magai, Consedine, Krivoshekova, Kudadjie-Gyamfi, & McPherson, 2006), this is the result of the emotion system becoming more complex and nuanced across the course of life, causing the acquisition of representational understanding and emotion regulatory capacities as we grow older. This way, older adults are more capable to cope with constraints of later life and with the negative events that are more likely to take place (Pasupathi et al., 1998). Over the years, alternative theories of the positivity effect have been suggested. For example, Labouvie-Vief, Grühn, and Studer (2010) stated in their dynamic integration theory that age-related cognitive declines lead to automatic processing of positive information, because it is easier to process than negative information. However, perspectives like this are inconsistent with important empirical findings that support the evidence in favor of the positivity effect, namely in visual stimuli (e.g., Spaniol, Voss, & Grady, 2008) and lexical stimuli (e.g., Shamaskin, Mikels, & Reed, 2010).

The experience of emotions in older adults has been broadly studied, as outlined above. The speech *perception* of emotional expression in older adults has been investigated previously, too (e.g., Schmidt, Janse, & Scharenborg, 2016), but the speech *production* of emotional expression in older adults has received much less attention. Because of the previously mentioned positivity effect, it cannot be assumed that results on speech production of emotional expression in *younger* adults can be generalized to *older* adults. They experience emotions in a different way than younger adults, so it is likely their emotional speech production is influenced by their age as well. For example, as multiple studies showed (Eichhorn, Kent, Austin, & Vorperian, 2017; Linville, 2001; Reubold, Harrington, & Kleber, 2010), the impact of aging on F_0 affects both women and men. In female aging voices, F_0 decreases approximately 30 Hz around the menopause. The male voice follows a pattern of a decrease followed by a strong increase of approximately 30 Hz, starting around the fiftieth year. In addition, intensity and tempo features are affected by age as well. Maximum vowel intensity and articulation rate decrease in both women and men with increasing age (Linville, 2001). The effect of age on voice quality features seems to vary from one speaker to another. Overall, more research has to be conducted to comprehend the nuances in spontaneous affective speech production of older adults.

2.2 Autobiographical memory

2.2.1 Emotional autobiographical memories

The previously mentioned positivity effect is not the only emotional difference between younger and older adults. It also seems that the autobiographical memory, the totality of memories of one's own life (Meeter & Hendriks, 2012), is subject to a valence-related bias. Generally, findings suggest that the valence of recalled memories is congruent with the current mood state (e.g., Matt, Vasquez, & Campbell, 1992). However, there are also examples of studies that showed the recall of positive memories while participants found themselves in a more negative mood (e.g., Rusting, 1998). These inconsistencies could be explained by the changing subjective experience (also called 'phenomenology') of autobiographical memory as an individual's life story evolves across adulthood (Luchetti & Sutin, 2018). Autobiographical memories are memories from one's own personal past that can be of prior experiences, facts about their own lives, or encounters with other people (Luchetti & Sutin, 2018; Xu et al., 2018), such as the birth of a child, the start of a new job, or the loss of a parent. Some memories are experienced as vivid, emotionally intense, and very detailed, whereas others are vague and fragmented. The very clear and specific memories are often 'self-defining' moments that can have a long-term impact on the identity of the individual, and early adulthood in particular consists of many moments like this (Conway, 2005). Especially for older adults the preservation of these memories remains important, as meaningful memories contribute to the individual's sense of overall well-being (e.g., Habermas & Köber, 2015). This is also confirmed in studies with individuals with depressive disorders that have more difficulty retrieving specific autobiographical memories than nondepressed individuals. For example, Williams (1996) asked participants with major depressive disorder (MDD) to retrieve a specific memory to multiple cues (e.g., *happy*, *alone*), using the Autobiographical Memory Task (AMT; Williams & Broadbent, 1986). Compared to controls, depressed individuals responded relatively more often with common memories that summarize across categories of similar events (e.g., "when it is my birthday", rather than "on my eighteenth birthday when my father bought me a brand-new car"). This reduced autobiographical memory specificity has shown to be a vulnerability factor for depression (Raes et al., 2006).

2.2.2 The reminiscence bump

For older adults it seems that the earlier mentioned self-defining moments that remain vivid in the mind (Luchetti & Sutin, 2018) are related to a phenomenon called the reminiscence bump. This means that people above the age of 40 tend to recall significantly more memories from their adolescence and early adulthood (15-30 years of age) than from other periods in their lives. The reminiscence bump only appears to exist for positive memories and not for negative memories (Berntsen & Rubin, 2002; 2004; Glück & Bluck, 2007; Rubin & Berntsen, 2003). For example, Berntsen and Rubin (2002) found that the reminiscence bump appeared when people were asked for their happiest and most important memories, but not when they were asked for their saddest and most traumatic memories. This can be explained by the proposal of Berntsen and Rubin (2002; 2004; Rubin & Berntsen, 2003) of a cultural life script theory, which suggests that the recall of autobiographical memories is guided by an underlying cultural life script. Life scripts can be described as cognitive schemes of people in a given culture regarding what transitional events a typical individual is likely to experience during the life course, as well as the age at which these events are likely to be experienced (Erdoğan, Baran, Avlar, Taş, & Tekcan, 2008). According to this, life scripts serve as guides for the retrieval of autobiographical memories.

2.2.3 Cultural life scripts

Berntsen and Rubin (2004) were the first to examine the positivity bias in the reminiscence bump. They asked 103 undergraduates to imagine a prototypical infant and to name the seven most important events that were likely to take place in the life of this person. Participants then specified at what age these events were most likely to occur and rated the events on prevalence, importance, and valence. The reported events were mainly normative life events, such as graduation, marriage, and having children, and the majority of the events was positive, which suggested that life scripts represent an idealized and not an ordinary life, from which many common and some important events can be left out (Berntsen & Rubin, 2004; Rubin, Berntsen, & Hutson, 2009). Furthermore, most positive events were expected to happen during early adulthood, whereas the negative events could occur at any time in life. Also, the study's findings could not be attributed to the participants' own life experiences, because they had not experienced most of the reported events yet, like retirement, getting grandchildren, or the death of a partner, which means that a life script does not contain personal memories of the particular life events (Schank & Abelson, 1977). These results were then replicated, confirmed, and extended by many additional studies, a portion of which will be discussed next.

One of the first to replicate the study of Berntsen and Rubin (2004) were Erdoğan and colleagues (2008). They added one significant component to the research design: instead of asking all 200 participants to imagine a prototypical newborn infant when listing the seven most important events that were most likely to occur in his or her life, they asked half of the participants to imagine an older adult of 90 years of age and to list the seven most important events that had already taken place in his or her life. The reported events of both imagined persons were then rated on prevalence, importance, and valence, and participants specified the age at which the events were most likely to (have) happen(ed). By adding a prototypical older person to the design, Erdoğan and colleagues (2008) investigated how the age of the target person affected the life scripts themselves and the valence of the life scripts. They concluded that the life scripts for their samples overlapped substantially with earlier data from Berntsen and Rubin (2004), and that the reported events (and their valence) were not influenced by the age of the target person.

Janssen and Rubin (2011) replicated the original study with the purpose to confirm the suggestion that life scripts are cultural semantic knowledge and therefore should be known by all adult age groups, including those who have not lived through all events in the life script. For this reason, they asked three groups (595 participants in total) of young (16-35 years of age), middle-aged (36-55 years of age), and older adults (56-75 years of age) to imagine an ordinary, prototypically infant and to name the seven most important events that were likely to take place in his or her life. They too were subsequently asked to answer questions about at what age these events were expected to occur and about their prevalence, importance, and valence. Janssen and Rubin (2011) concluded that the cultural life scripts were indeed the same for all age groups, which is in line with the assumption that people who have lived through a smaller part of their life still know the entire life script of their culture. They are not extracted from personal memories or experiences, but are transmitted by tradition.

Additionally, Janssen (2015) examined if cultural life scripts also existed for public events, like the death of a famous person, royal events, terrorist attacks, or environmental disasters. He was motivated by findings of the reminiscence bump also being present in the memory of public events (e.g., Janssen, Murre, & Meeter, 2008). To examine this, half of the in total 209 participants were asked to follow the original life script procedure as proposed by Berntsen and Rubin (2004), while the other half of the participants were asked to list the seven most important *public* events that will most likely take place during the life of a prototypical infant. Both groups were asked to rate the events on prevalence (or likelihood of occurrence for public events), importance, age, and valence. Janssen (2015) found no support for cultural life scripts as an explanation for the reminiscence bump in the memory of public events; most public events were expected to occur before the reminiscence bump. Although there was some agreement on which public events were likely to happen in a prototypical person's life (such as a sports event, elections, a medical breakthrough, or war), there was little agreement on *when* these events were supposed to occur.

Rubin and colleagues (2009) replicated the original study with the additional intention to investigate individual differences in the life script and life story. The further a person's personal life story is from the normative life script, the more likely it is they are less well tuned to the expectations of their culture and may have more personal and emotional difficulties (Rubin et al., 2009). In order to examine this, 100 participants were asked to generate seven events that would go into the life script and their own life story, respectively. They were asked to rate the events on prevalence, importance, age, and valence. These measures were then compared to measures of depression under the expectation that people who are more depressed will have events and autobiographical memories of a low valence more easily available (Williams, 1996), and will report them among the events of a life script as well as their own life story. The measures were also compared to an inventory of post-traumatic stress disorder (PTSD) symptoms, because it was expected that having a highly negative event as easily available is more likely to lead to a view of one's own life that is less prototypical, which means that the life story and life script are likely to not match. Rubin and colleagues (2009) found that the results were comparable to the results found by Berntsen and Rubin (2004), and that the valence of life story events indeed correlated with life script valence, depression, and PTSD symptoms.

Grysman and Dimakis (2017) used the same methodology as Berntsen and Rubin (2004), but with a small though important adjustment. They wanted to examine the expectations of older adults for life events occurring in middle and later adulthood (and thus after the reminiscence bump) and if such later life events are scripted as well. If these life events were to be scripted too, Berntsen and Rubin (2004) made six predictions, among which that scripted events had to be shared by many people, include predominantly positive events, and had to be dominated by “culturally sanctioned transitional events” and not by biological events, such as *partner’s death* or *serious disease*. In Grysman and Dimakis’ (2017) study, 100 participants took part who ranged in age from 38 to 76. They were asked to name seven events they expected to occur in the life course of a prototypical person of their own age. Subsequently, they stated the prevalence, importance, age, and valence of the reported events. The participants were also asked if they expected their own personal life to follow a similar path as the seven events they just outlined. Ultimately, Grysman and Dimakis (2017) concluded that the reported events were in fact scripted and of a high valence, despite the events occurring after the highly positive reminiscence bump.

Overall, all findings of the studies described above were in line with the proposal that life events are merely positive events and often occur between the ages of 15 to 30. When asked to imagine a prototypical older adult of their own age, participants still reported mostly positive events that also seem to be scripted (Erdoğan et al., 2008; Grysman & Dimakis, 2017). An overview of the life events (and their valence scores) reported by the participants of the studies described above, is shown in Table 1. Here, the life events are sorted from low to high, based on their valence score. The results of both target persons of Erdoğan and colleagues’ study (2008) are adopted in Table 1, *Erdoğan et al. (i)* representing the results of the prototypical newborn infant, and *Erdoğan et al. (o)* representing the results of the prototypical older adult. The life events of Janssen represented in Table 1 are the results from the personal life scripts, and not of the public events. As can be seen, some life events are well represented in the studies portrayed above, such as *own marriage*, *falling in love*, and *child’s birth*. Other events, with a notably lower valence, are less common; for example *sibling’s death*, *neglected by children*, and *operation*. This is yet another piece of evidence that people expect that mostly positive events are likely to occur in a prototypical life. The results shown in Table 1 will lay the basis for a newly developed valence scheme for annotating spontaneous emotions, which will fulfill the need for a better way to measure emotions in spontaneous affective speech.

2.3 Lexical speech features of emotion

The cultural life script studies described above and the outcomes in Table 1 show that it is possible to quantify lexical speech features of emotion. Lexical features refer to the content of what is said or the word usage in affective speech, without taking any acoustic (such as pitch or intensity) or nonlinguistic (such as crying or laughing) features into consideration. Studies of lexical features of emotion have aimed to describe affective key words in terms of emotion dimensions. For instance, Moors and colleagues (2013) defined the norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. The set of words consisted mainly of nouns, adjectives, adverbs, and verbs. The affective ratings were performed by 224 students, who were subdivided in equally sized groups of males and females from two Belgian and two Dutch samples. One of the strengths of this study is that each participant rated the entire set of words for only one affective variable (valence, arousal, dominance, or age of acquisition). This way, the ratings for one variable could not be influenced by the ratings for another variable. In each sample, each variable was rated by 16 students. Participants in the valence condition were asked to judge the extent to which the words referred to something that is positive/pleasant or

Table 1. *Overview of mean valence scores, extracted from Berntsen & Rubin (2004), Erdoğan et al. (2008), Grysman & Dimakis (2017), Janssen (2015), Janssen & Rubin (2011), and Rubin et al. (2009). Valence score: 1 = low, 7 = high.*

Life event	**Valence**							
	Berntsen & Rubin	Erdoğan et al. (i)	Erdoğan et al. (o)	Grysman & Dimakis	Janssen	Janssen & Rubin	Rubin et al.	**M**
Child's death	-	-	1.14	-	-	-	-	**1.14**
Partner's death	1.00	-	1.04	1.13	1.00	1.67	-	**1.17**
Sibling's death	-	-	-	-	1.33	-	-	**1.33**
Parent's death	-	1.50	1.30	-	1.09	2.03	1.10	**1.40**
Neglected by children	-	-	1.50	-	-	-	-	**1.50**
War	-	-	1.83	-	1.29	-	-	**1.56**
Grandparent's death	-	-	-	-	1.00	2.19	-	**1.60**
First rejection	1.00	-	-	-	1.75	2.48	-	**1.74**
Friend's death	-	-	-	-	1.33	2.19	-	**1.76**
Serious disease	2.67	1.40	1.85	1.69	-	2.00	-	**1.92**
Own divorce	2.00	2.00	1.75	2.09	-	2.14	-	**2.00**
Career failure	-	-	2.00	-	-	2.00	-	**2.00**
Parents' divorce	-	-	-	-	2.00	2.15	-	**2.08**
Infidelity	-	-	2.12	-	-	-	-	**2.12**
Financial troubles	-	-	3.00	1.42	-	-	-	**2.21**
In an accident	-	1.86	2.00	-	3.00	2.35	-	**2.30**
Operation	-	-	2.40	-	-	-	-	**2.40**
Family quarrels	-	2.50	2.40	-	-	-	-	**2.45**
Caring for parents	-	-	-	2.75	-	-	-	**2.75**
Psychological problems	-	-	3.00	-	-	-	-	**3.00**
Empty nest	3.25	-	-	3.80	-	5.00	3.25	**3.83**
Prepare for death	-	-	-	4.13	-	-	-	**4.13**
Move	-	2.75	3.80	4.80	4.50	4.80	-	**4.13**
Other child's milestones	-	-	-	4.20	-	-	-	**4.20**
Puberty	3.82	4.06	5.00	-	6.40	4.82	4.00	**4.68**
Military service	-	4.91	4.78	-	-	-	-	**4.85**
Retirement	3.94	4.70	5.17	5.49	5.09	4.86	5.55	**4.97**
Leave home	5.12	-	4.20	-	5.44	5.73	5.63	**5.22**
Sibling's birth	4.58	5.33	-	-	6.33	5.34	-	**5.40**
First job	5.00	5.51	5.82	-	-	6.00	5.06	**5.48**
Begin school	5.24	5.77	5.80	-	-	5.65	5.19	**5.53**
High school	-	5.78	5.25	-	5.20	5.44	6.09	**5.55**
First sexual experience	5.50	-	6.00	-	5.50	5.65	5.56	**5.64**
College	5.30	6.30	5.96	-	5.59	5.95	6.17	**5.88**
Career success	-	-	-	6.25	5.75	5.82	-	**5.94**
Buying a house	-	-	6.28	-	6.15	5.76	-	**6.06**
Travelling	5.30	-	-	6.42	6.80	6.04	-	**6.14**
Driver's license	-	-	-	-	6.63	6.00	6.08	**6.24**
Big achievement	6.00	-	-	-	6.75	6.10	-	**6.28**
Having friends	6.60	-	-	-	6.33	5.97	-	**6.30**
Own marriage	6.52	5.99	5.88	6.29	6.57	6.35	6.70	**6.33**
Falling in love	6.44	5.92	6.19	7.00	6.45	6.00	6.36	**6.34**
Child's birth	6.58	6.55	6.64	5.00	6.25	6.60	6.74	**6.34**
The 'right' job	7.00	-	6.30	-	-	5.80	-	**6.37**

Table 1. *Continued.*

Life event	**Valence**							
	Berntsen & Rubin	Erdoğan et al. (i)	Erdoğan et al. (o)	Grysman & Dimakis	Janssen	Janssen & Rubin	Rubin et al.	**M**
Child's marriage	-	-	6.33	6.60	-	-	-	**6.47**
Graduation	-	-	-	-	6.50	6.57	-	**6.54**
Child's college grad.	-	-	6.71	6.60	-	-	-	**6.66**
Grandchild's birth	6.73	-	6.72	6.94	6.60	6.40	6.54	**6.66**
Family holidays	-	-	-	-	6.67	-	-	**6.67**
Celebrations	-	-	-	6.75	-	-	-	**6.75**

negative/unpleasant, using a 7-point Likert scale (1 = very negative/unpleasant, 7 = very positive/pleasant). Moors and colleagues (2013) found high correlations within each group of raters and high correlations between the groups of raters for each variable. Most importantly, they found that the valence ratings of a previous study (Hermans & De Houwer, 1994) correlated highly with those of their study. This means that valence can also be determined by looking at key words in text or the word usage in speech.

The drawback of applying Moors and colleagues' (2013) method to a speech sample is that certain word combinations could get incorrect valence ratings. For example, *vreselijk mooi*, meaning *terribly beautiful*, would not be rated like an intensive form of *beautiful*, but as *terrible + beautiful* with isolated negative and positive valence ratings. A study that anticipated on this issue is that of De Smedt and Daelemans (2012b). They introduced another way to determine the valence of lexical features. They presented a new open source subjectivity lexicon for Dutch adjectives. The lexicon is a dictionary of 1,100 adjectives that are manually annotated by seven human annotators with polarity strength, subjectivity, and intensity for each word sense. Polarity can be classified into negative, neutral, and positive valence (Moors et al., 2013). The created lexicon is part of PATTERN (De Smedt & Daelemans, 2012a), an open source lexicon with an algorithm that applies sense discrimination by taking into account intensifiers, downtoners, and negations. Downtoners strengthen or diminish the sentiment of an adjective by using an adverb (e.g., *ongelooflijk goed*, meaning *incredibly good*). Negations provide the distinction of PATTERN between, for example, *echt niet blij* and *niet echt blij*, meaning *really not happy* and *not really happy*, respectively. This sentiment analysis aims at the determination of polarity of text and is thus suitable for establishing the valence of word usage in emotional speech production.

2.4 Acoustic speech features of emotion

As briefly described in section 2.1.4, many studies of vocal expression have attempted to specify which acoustic features of speech are able to characterize affective speech production. According to the source filter model of speech production (Fant, 1960), different voice cues can be measured to provide information about emotional states. However, this has proved to be more difficult than expected. This could be explained by the matter that most vocal expression studies have focused on discrete emotions. It has been suggested that voice changes can be associated with specific emotions (e.g., Juslin & Laukka, 2003); for example, a strong and rough voice could be associated with anger (Gobl & Chasaide, 2003). However, it seems that the affective states expressed most often in spontaneous speech rather are milder forms of affective states (Laukka et al., 2005), that can better be described in terms of emotion dimensions like arousal and valence. As stated before, arousal seems to have a great and consistent influence on vocal expression (Banse & Scherer, 1996). For instance, it seems that high arousal can be associated with a high mean F_0, a large F_0 variability, a fast speech rate, short pauses, increased voice intensity, and increased high-frequency energy (e.g., Breitenstein, Van Lancker, & Daum, 2001; Pereira, 2000). However, findings on valence are much more inconsistent (e.g., Bachorowski, 1999; Leinonen, Hiltunen, Linnankoski, & Laakso, 1997; Protopapas & Lieberman, 1997). Some studies have found that a positive valence can be characterized by a low mean F_0, a large F_0 variability, a fast speech rate, shorter pauses, and low voice intensity (e.g., Laukka et al., 2005; Scherer, 1972; Scherer & Oshinsky, 1977; Schröder et al., 2001), while other studies were not able to acquire any cues that explain different levels of valence (see Juslin & Scherer, 2005).

A concrete example comes from Scherer and Oshinsky (1977), who studied the emotional expression in speech and music by letting participants judge several tones on multiple scales. Participants were 48 untrained raters who were exposed to three types of tone sequences, consisting of eight tones. In total 188 stimuli were divided into four sets of 32 stimuli, that were presented to four rating groups, each group hearing only one stimulus set. Each participant rated each tone sequence on three 10-point scales: a valence, arousal, and potency scale. Each participant also indicated whether a sequence could or could not be an expression of one of seven emotions, namely anger, fear, boredom, surprise, happiness, sadness, and disgust. Results showed that positive valence could generally be associated with a downward slope of the pitch contour, but also with the combination of large pitch variation and an upward slope of the pitch contour. Scherer and Oshinsky (1977) stated that this contradiction could possibly be explained by the presence of different types of happiness that are communicated by specific configurations of acoustic cues (i.e., quiet satisfaction versus cheerful celebration).

In contrast to Scherer and Oshinsky (1977), Schröder and colleagues (2001) worked with actual speech data and investigated the relationship between three emotion dimensions (arousal, valence, and power) and multiple acoustic features relevant for speech synthesis. The natural speech data consisted of a database of TV recordings of chat shows, religious programs, and interviews recorded in a studio. Participants were asked to locate the emotional tone of a recording in a two-dimensional arousal-valence space, continuously over time. The power dimension was measured in a different way. The acoustic analysis of the database consisted of 26 acoustic variables in total, among which variables for intonation (e.g., F_0), speech rate (e.g., duration of pauses), intensity (e.g., mean intensity and intensity range), and voice quality (e.g., spectral slope). The most fundamental result was that nearly all acoustic features significantly correlated with the emotion dimensions. Correlations with the valence dimension were less strong than the arousal and power dimensions, but they were systematic. Negative valence was associated with longer pauses, faster F_0 drops, increased intensity, and more prominent intensity maxima, which is in contrast with the results of Scherer and Oshinsky (1977).

Another example comes from Laukka and colleagues (2005), who aimed to explore whether specific acoustic features are associated with specific emotion dimensions. Eight professional actors vocally portrayed five different emotions (anger, fear, disgust, happiness, and sadness) with both weak and strong emotion intensity by reading short phrases aloud (i.e., "It is eleven o'clock"). The actors were also instructed to perform the same material without any expression. The emotion portrayals were analyzed on 20 acoustic features, among which F_0, voice intensity, the first three formants, and speech rate. Then, one group of 30 students and one group of six expert judges took part in the listening experiments. All participants were instructed to listen to all emotion portrayals of the actors and rate the portrayals on four emotional dimension scales, namely arousal, valence, potency, and intensity. The ratings were made on scales ranging from 0 (low arousal, negative valence, low potency, and low intensity) to 10 (high arousal, positive valence, high potency, and high intensity). The results showed that each of the four dimensions was significantly correlated with a number of acoustic features. Positive valence was associated with a low mean F_0, low minimum F_0, low mean voice intensity, small voice intensity variability, fast speech rate, low first formant, and little high-frequency energy (cut-offs at both 500 and 1000 Hz). Furthermore, Laukka and colleagues

(2005) concluded that the listeners' mean ratings could be successfully predicted from the acoustic features for all dimensions, except valence. An explanation could be that acoustic features associated with valence are more independent of autonomic physiological changes and that valence perception may become better differentiated as the length of the spoken phrase increases. This suggests that valence could be associated with other features, like speech rhythm.

Tahon and colleagues (2012) examined the relevance of acoustic features for valence detection. They wanted to improve an already existing valence detection system by testing new voice quality features. In contrast to Laukka and colleagues (2005), they did not use actors for the emotional speech data, but 22 older people who behaved as they would probably do in everyday life. Their speech was evoked by interacting with a social humanoid robot that is able to assist people at home in everyday activities. The speakers were asked to imagine multiple scenarios in which they pictured themselves in a situation of waking up in the morning. The robot would come to them to chat about, for example, their health or their plans of the day. To each scenario one affective state was applied, which the speaker was asked by the robot to imagine: well-being, minor illness, depressed, medical distress, or happy. For each scenario, the robot had a different social attitude: positive (friendly, empathetic, and encouraging) or negative (directive, doubtful, and machine-like). The six acoustic features used for valence detection were voicing, harmonics-to-noise ratio (the proportion of noise in the speech signal), jitter and shimmer (to evaluate the small time variation of F_0 and energy), and two new voice cues: the relaxation coefficient (Rd) and the functions of phase-distortion (FPD). The Rd coefficient is a parameter that indicates the relaxation of the voice; the stronger Rd is, the more relaxed the voice is (0 = very tense, 2.5 = relaxed). The FPD indicate the distortion of the phase spectrum around its linear phase component. For both parameters, the mean value and standard deviations are reported. The results showed that only four of the six voice quality features were interesting for valence discrimination: harmonics-to-noise ratio (HNR), unvoiced ratio, Rd (mean and standard deviation), and FPD. Specifically, FPD associated with F_0 features and shimmer associated with F_0 and energy features were interesting for valence detection. Although Rd and FPD features were only computed on voiced parts of the speech data, these new voice quality features could be useful for emotion detection, too. Further research is necessary, however, to ensure the relevance of these acoustic features for valence detection.

One study with promising results is that of Goudbeek and Scherer (2010). They examined the role of emotion dimensions other than arousal with the newly developed Geneva Multimodal Emotion Portrayals corpus, a corpus that contains 12 emotions that systematically vary with respect to valence, arousal, and potency/control. The 12 emotions were elation (joy), amusement, pride, pleasure, relief, interest, hot anger (rage), panic fear, despair, cold anger (irritation), anxiety, and sadness, and were portrayed by ten professional actors. The emotions were expressed in two meaningless carrier sentences (for example, "ne kali bam soud molen", uttered with declarative prosody). In addition, the actors were asked to express the emotion by only using the sustained vowel /a/. Of the complete utterances of the actors, 26 acoustic features were extracted, among which duration (of all utterances, of the (un)voiced and silent parts), speech rate, various F_0 and intensity parameters, energy, and spectral slope. After applying regression analyses to investigate the effect of arousal, valence, and potency/control on seven vocal parameters, it seemed that at low levels of arousal, positive valence was reflected in lower intensity variability and a steeper spectral slope than negative valence. At high levels of arousal, positive valence was associated with a lower level of intensity, more variation in intensity, a less noisy signal and a narrower spectrum than negative valence. Ultimately, Goudbeek and Scherer (2010) concluded that although arousal still dominated many acoustic features, it is possible to identify features that are specifically related to potency/control and, most important, valence.

Table 2 shows an overview of the studies discussed above and various other studies that focused on the acoustic speech features of valence. These studies have all adopted the dimensional approach to search for acoustic correlates of various emotional dimensions. Overall, to obtain valence differentiation, it seems necessary to reach beyond single measures of the most common acoustic features, such as F_0, speech rate, and intensity, and to analyze other cues that differentiate among emotions (Juslin & Scherer, 2005). Voice quality parameters, like spectral slope, and intensity parameters seem most promising for the detection of valence. Combining lexical and acoustic features could hold the key to establishing the valence of spontaneous affective speech production.

2.5 Research questions and hypotheses

As discussed above, not all dimensions of emotion are equally examined in affective speech research. Arousal has clear and consistent correlations with acoustic features, in contrast to valence – not to mention other possible emotion dimensions (Fontaine et al., 2007; Goudbeek & Scherer, 2010; Laukka et al., 2005, Schröder et al., 2001). Furthermore, studies that did look at the valence of speech production often did not include the speech of older adults (≥ 65 years of age). Since measuring the valence of spontaneous affective speech production seems challenging, the aims of this study are to find a proper way to measure valence and to look at the relations between lexical and acoustic speech features. Therefore, the following general research question is addressed in the current study:

> *How can valence be measured in spontaneous emotional speech production of healthy older adults (≥ 65 years of age)?*

To answer this research question properly four sub questions have been composed that, respectively, focus on the development of a novel valence scheme, the valence of lexical and acoustic features in the spoken emotional memories of healthy older adults, and the relation between lexical and acoustic speech features.

Since emotions and memories are highly subjective and personal, it is very difficult to obtain valence information using an objective measure. For this reason, the first sub question is as follows: *How can valence be annotated in a reliable manner in spontaneous spoken emotional memories?* With respect to this question, a novel annotation scheme is developed and proposed to measure the valence of the word usage in emotional autobiographical memories. This scheme does not only determine the valence of a certain memory, but also allows for subjective interpretations, which has not yet been seen in affective research before. In section 3.2.2, this novel valence scheme will be discussed in further detail.

The second sub question focuses on the valence of word usage in spontaneous speech: *To what extent can sentiment analysis predict valence in spontaneous spoken emotional memories?* The word usage in emotional speech will be measured with the sentiment analysis of De Smedt and Daelemans (2012b). The expectation is that the results of the sentiment analysis are positively associated with the valence scores obtained by the newly developed annotation scheme, since the valence scores and the sentiment analysis are both based on the lexical features of speech.

The third sub question concerns the valence of the acoustic features in spontaneous speech: *To what extent can acoustic speech features predict valence in spontaneous spoken emotional memories?* Expectations are that not all acoustic features can predict valence, but it is expected that parameters of voice quality (i.e., spectral slope) and intensity (Goudbeek & Scherer, 2010; Tahon et al., 2012; Schröder et al., 2001) are positively associated with valence.

Table 2. *Elicitation methods of various studies on acoustic speech features of emotion.*

Study	**Speech samples**	**Dimensions studied**	**Results on acoustic correlates of valence**
Busso & Rahman (2012)[2]	Corpus recordings	Valence	Lower F_0 median, small F_0 values, high spectral feature values[3]
Goudbeek & Scherer (2010)[1]	Professional actors	Valence, arousal, potency	Lower intensity variability, steeper spectral slope, lower levels of intensity, more intensity variation, less noisy signal, narrow spectrum[3]
Laukka et al. (2005)[1]	Professional actors	Valence, arousal, potency, intensity	Low mean F_0, low minimum F_0, low mean intensity, small intensity variability, fast speech rate, low F_1, little high-frequency energy, F_1 precision[3]
Scherer (1972)[2]	Nonverbal tone sequences	Valence, arousal, potency	Moderate F_0 variation[4], extreme F_0 variation[3]
Scherer & Oshinsky (1977)[1]	Nonverbal tone sequences	Valence, arousal, potency	Decline of F_0 contour, combination of large F_0 variation and increase of F_0 contour[3]
Schröder et al. (2001)[1]	TV and interview recordings	Valence, arousal, power	Longer pauses, faster F_0 drops, increased intensity, more prominent intensity maxima[4]
Tahon et al. (2012)[1]	Spontaneous speech elicited by a humanoid robot	Valence	HNR, unvoiced ratio, Rd (mean and standard deviation), and FPD are interesting for *detecting* valence

[1]Studies extensively described in the current study.
[2]Additional studies on acoustic correlates of emotional valence.
[3]Associations with *positive* valence.
[4]Associations with *negative* valence.

Parameters such as F_0 and speech rate seem less promising for the detection of valence in spontaneous affective speech (Juslin & Scherer, 2005; Laukka et al., 2005, Schröder et al., 2001).

The fourth and last sub question addresses the relation between the valence of lexical and acoustic speech features in spontaneous affective speech: *How are automated analyses of lexical and acoustic features in speech production related to each other with respect to valence expression in spontaneous spoken emotional memories?* This question essentially investigates if the way something is said (acoustic features) compares to what has been said (lexical features). Even though there are no previous studies that accounted for this expectation, it is assumed that the valence of lexical and acoustic features of speech production match. It is most likely that high lexical valence values are related to high acoustic valence values.

3. Methods

In this chapter, the methods of the current study will be outlined. First, the participant's demographics will be discussed. Second, the materials for gathering the data will be described. In this section, a newly developed scheme for annotating the valence of emotional memories will be presented. Then, the analyses for the lexical and acoustic speech features will be discussed, and lastly, a brief overview of the data analysis will be given.

3.1 Participants

This study is part of a larger project on vocal expressions of spontaneous affect in older adults with mild dementia. To that end, data has been collected with vocal expressions of spontaneous affect in healthy older adults. The current study focuses on a subset of these data. In total, there were 23 participants who took part in the study. In the current study, the data of 12 participants (seven males, five females) is analyzed. These 12 participants were aged between 66 and 81 years old (M = 73.00, SD = 5.26). An overview of the participants' demographics can be found in Appendix I. The participants were recruited through advertisements in local newspapers. In order to engage in the study, participants had to be at least 65 years of age, have unimpaired or corrected vision and hearing, and had to speak and read fluently Dutch. Participants were excluded based on memory problems, traumatic experiences, and having a pacemaker, the latter because of a follow-up meeting of the larger project in which the physiology in participants was measured. If the participants agreed to take part in the study, they received a more detailed information letter (see Appendix II). The data was collected through interviews that took place at the participant's home or a location where the participant felt comfortable.

3.2 Materials

3.2.1 Autobiographical Memory Test (AMT)

Data was collected using a revised version of the Autobiographical Memory Task (AMT; Williams & Broadbent, 1986). The AMT served as a word association task to elicit spontaneous emotional memories. Two emotional cue words of positive and negative valence, namely *happy* and *sad*, were presented, after which three specific memories for each cue word had to be described by the participants, giving a total of (at least) six memories per participant. A specific memory is a particular personal event in someone's life that happened only once on a certain time and day and did not last longer than one day (Raes et al., 2006). Beforehand, two neutral cue words (*grass* and *bread*) were used to practice the retrieval of memories. The fixed order of all cue words was *grass*, *bread*, *sad*, and *happy*.

Three microphones were used to record the interviews. Wireless lavalier microphones were placed around the necks of the participant and the interviewer and one shotgun microphone was placed between the participant and the interviewer. Only the recordings of the wireless lavalier microphones that were placed around the neck of the participants were used for the analyses.

3.2.2 Valence of Emotional Memories scale (VEM)

Since existing valence annotating schemes (e.g., Aubergé et al., 2006; Kanluan et al., 2008) did not account for the complexity of emotional memories, a novel coding scheme was developed for the establishment of the valence of emotional autobiographical memories. The Valence of Emotional Memories scale (VEM) is based on recent research on the valence of life events (Berntsen & Rubin, 2004; Erdoğan et al., 2008; Grysman & Dimakis, 2017; Janssen, 2015; Janssen & Rubin, 2011; Rubin et al., 2009) and separate Dutch key words (Moors et al., 2013). The VEM consists of 53 life events with corresponding valence scores between 1 (low) and 7 (high)[2]. Three of the life events represented in the VEM were not mentioned in previous research. The valence of two of these life events, *hobby* and *youth*, was determined by the mean valence scores of multiple Dutch key words based on the research of Moors and colleagues (2013). Two raters selected the key words to represent these life events and the corresponding valence scores (see Table 3). The valence of the remaining life event, *grandchild's death*, was created due to the mention of this specific life event across participants. The valence score of *grandchild's death* was determined in agreement of two raters as well, based on the same valence score of *child's death* (valence score = 1.14).

Table 3. *Mean valence of various hobbies (left) and mean valence of various key words about 'youth' (right), all extracted from Moors et al. (2013).*

Key word	**Valence**	**Key word**	**Valence**
To paint	4.30	Toddler (*kleuter*)	4.56
To walk	4.36	Toddler (*peuter*)	4.58
To bake	4.78	Cub	4.87
To cook	4.86	Child	5.05
To bike	4.88	Innocence	5.09
To skate	4.89	Youth	5.25
To sing	5.22	Young	5.30
To dance	5.56	To grow	5.31
Mean	**4.86**	**Mean**	**5.00**

Based on Tables 1 and 3, the definitive VEM was created (see Table 4). On the left of Table 4, specific events (*specifics*) are arranged on the basis of a common factor (*life event*). For example, *birth* is the life event and can be specified by *sibling*, *child*, or *grandchild*. On the right, the same table is presented, but sorted from low to high valence.

All emotional memories gathered in the course of the AMT were transcribed (first automatically, using a Dutch open source speech recognizer trained with Kaldi (Povey et al., 2011, see https://github.com/opensource-spraakherkenning-nl/Kaldi_NL), and afterwards checked and if needed manually adjusted), anonymized, and chunked into meaningful fragments based on the content by two independent raters. Consequently, one memory could contain more than one fragment, separated by a turning point or new theme. For example, a sad memory about the loss of a loved one could be divided into several life events, such as *travelling*, *serious disease*, and eventually *death* within the same memory. In total, all emotional memories were chunked into 234 fragments. Then, two independent raters evaluated the transcripts of the fragments on valence.

Two independent raters were presented with the transcripts of the audio recordings of the spoken memories of the participants. The raters received specific instructions for annotating the valence of the fragments (see Appendix III). The valence score is determined by establishing the valence of a particular life event (and if relevant, its specifics), a subjectivity score, and/or a reflective score. A subjectivity score of -1/+1 is only applied if the participant explicitly states that his/her memory is different in terms of valence than the VEM describes. For instance, if a participant states the following example in (1), the valence score for *death* > *parent* is selected and a subjectivity score of +1 is added, resulting in a total valence score of 1.40 + 1 = 2.40.

[2] Originally, studies on the valence of life events expressed valence in scores between -3 (low) and +3 (high). Since Moors and colleagues (2013) used scores between 1 and 7, all valence scores of the VEM were converted to the same classification system. The choice fell on the classification system of Moors and colleagues (2013), because their lexicon could be used to create new life events (and their corresponding valence scores) that are not represented in the VEM as shown in Table 4.

Table 4. *Valence of Emotional Memories scale (VEM). Left: specifics are arranged on the basis of a common factor. Right: life events are sorted from low to high valence score.*

Life event	Specifics	Valence
Death	Child	1.14
	Grandchild	1.14
	Partner	1.17
	Sibling	1.33
	Parent	1.40
	Grandparent	1.60
	Friend	1.76
Neglected by children		1.50
War		1.56
First rejection		1.74
Serious disease		1.92
Divorce	Own	2.00
	Parents	2.08
Career failure		2.00
Infidelity		2.12
Financial troubles		2.21
In an accident		2.30
Operation		2.40
Family quarrels		2.45
Caring for parents		2.75
Psychological problems		3.00
Empty nest		3.83
Prepare for death		4.13
Move		4.13
Other child's milestones		4.20
Puberty		4.68
Military service		4.85
Hobby		4.86
Retirement		4.97
Youth		5.00
Leave home		5.22
Begin school		5.53
High school		5.55
First sexual experience		5.64
College		5.88
Birth	Sibling	5.40
	Child	6.34
	Grandchild	6.66
Achievement	First job	5.48
	Career success	5.94
	Buying a house	6.06
	Driver's license	6.24
	Big achievement	6.28
	The 'right' job	6.37
	Graduation	6.54
Travelling		6.14
Having friends		6.30

Life event	Valence
Child's death	1.14
Grandchild's death	1.14
Partner's death	1.17
Sibling's death	1.33
Parent's death	1.40
Neglected by children	1.50
War	1.56
Grandparent's death	1.60
First rejection	1.74
Friend's death	1.76
Serious disease	1.92
Own divorce	2.00
Career failure	2.00
Parents' divorce	2.08
Infidelity	2.12
Financial troubles	2.21
In an accident	2.30
Operation	2.40
Family quarrels	2.45
Caring for parents	2.75
Psychological problems	3.00
Empty nest	3.83
Prepare for death	4.13
Move	4.13
Other child's milestones	4.20
Puberty	4.68
Military service	4.85
Hobby	4.86
Retirement	4.97
Youth	5.00
Leave home	5.22
Sibling's birth	5.40
First job	5.48
Begin school	5.53
High school	5.55
First sexual experience	5.64
College	5.88
Career success	5.94
Buying a house	6.06
Travelling	6.14
Driver's license	6.24
Big achievement	6.28
Having friends	6.30
Own marriage	6.33
Child's birth	6.34
Falling in love	6.34
The 'right' job	6.37

Table 4. *Continued.*

Life event	Specifics	Valence
Marriage	Own	6.33
	Child	6.47
Falling in love		6.34
Child's college graduation		6.66
Family holidays		6.67
Celebrations		6.75

Life event	Valence
Child's marriage	6.47
Graduation	6.54
Child's college graduation	6.66
Grandchild's birth	6.66
Family holidays	6.67
Celebrations	6.75

(1) *"Mijn moeder is overleden, maar die was 92 en dat zat er al twee jaar aan te komen. Natuurlijk speelt verdriet daar een rol bij, maar ook een stukje van 'het is beter zo.'* […] *Het was wel verdrietig, maar minder erg dan bij mijn vader."*
"My mother passed away, but she was 92 years old and it had been coming for two years. Of course, sadness plays a part in all of it, but there is also a feeling of acceptance. […] It was sad, but not as sad as when my father passed away."

The subjectivity score was created by the agreement of two independent raters in the need of a way to acknowledge and respect the highly subjective and personal emotional memories of the participants. It had been decided that a subjectivity score is defined as -1/+1 (and not a different score), because this way the positivity/negativity of the memory is still maintained. If, for example, a subjectivity score of -2/+2 was adopted, a negative life event such as *psychological problems* (with a valence of 3.00) could become a positive life event (3.00 + 2 = 5.00). With the current subjectivity scores, only a neutral life event such as *move* (with a valence of 4.13) can be changed into a positive/negative life event (for instance, 4.13 – 1 = 3.12). This is a completely new way of annotating valence and has not been seen in emotional research before. However, the possibility to manipulate the valence score in a constrained manner is necessary, because per definition emotions are not objective and it would be a lost cause trying to treat them as if they are. For the same reason, it is also possible to add life events to the VEM in the same way the life events *hobby* and *youth* were established based on the research of Moors and colleagues (2013). If needed, missing life events can be determined by selecting multiple relevant key words (and their valence scores) of the lexicon of Moors and colleagues (2013) and by calculating the mean valence score (see Table 3). This opportunity of adding new life events is of great importance, because life scripts (of which life events are derived) represent idealized and not ordinary lives, from which many common and some important events can be left out (Berntsen & Rubin, 2004; Rubin et al., 2009). Because of this, it is plausible that the VEM is not – and perhaps may never be – complete.

The valence scores of the participants' memories were first determined separately by the two independent raters and later discussed to establish intersubjective agreement. If a fragment did not contain a specific life event, but instead the participant reflected on his/her own emotions by explicitly describing emotions and/or reflections of their emotions, a reflective score is given by scoring the fragment a '1'. An example of such a memory is presented in example (2). Generally, no valence score is assigned to this type of memory. However, if the participant reflects on his/her emotions and describes a specific emotional memory in the same fragment, it is possible to assign both a valence score and a reflective score. The valence score is then the dominant score.

(2) *"Kijk, dat is een impact, maar de meeste verdrietige dingen… Ja, wat is verdrietig? Daar stap je zo weer overheen. En toevallig kom je dat een keer tegen, eventjes, en dan?* […] *Het begint iets meer te slijten, maar ik moet er niet over nadenken. Dan word je woest en verdrietig tegelijk."*

> "Look, it has a certain impact, but most sad things… Well, what is sadness? In time, you get over it. And sometimes it hits you, but then? […] It is slowly healing, but I don't want to think about it. It makes me furious and miserable at the same time."

The inter-rater agreement of the valence annotations of all 234 fragments of the emotional memories was calculated using Cohen's κ. There was moderate agreement (McHugh, 2012) between the two raters ($\kappa = .774$, $p < .001$). Out of the 234 memories, 205 were identified as life events. Figure 2 depicts the top ten of the most annotated life events of the VEM in these fragments. For example, it shows that 28 of the remaining 205 fragments were coded with the life event *youth* (13,7%).

Fragments were excluded from analysis if the valence scores established by the independent raters were considered to be neutral (valence scores between 3-5). Only fragments with valence scores between 1-3 (negative) and 5-7 (positive) were included, because the current study is just at the start of comprehending the roll of valence in older adult spontaneous affective speech. The next step in this process would be looking at all data, including neutral valence scores. After exclusion, 188 of the original 234 fragments remained for further analysis.

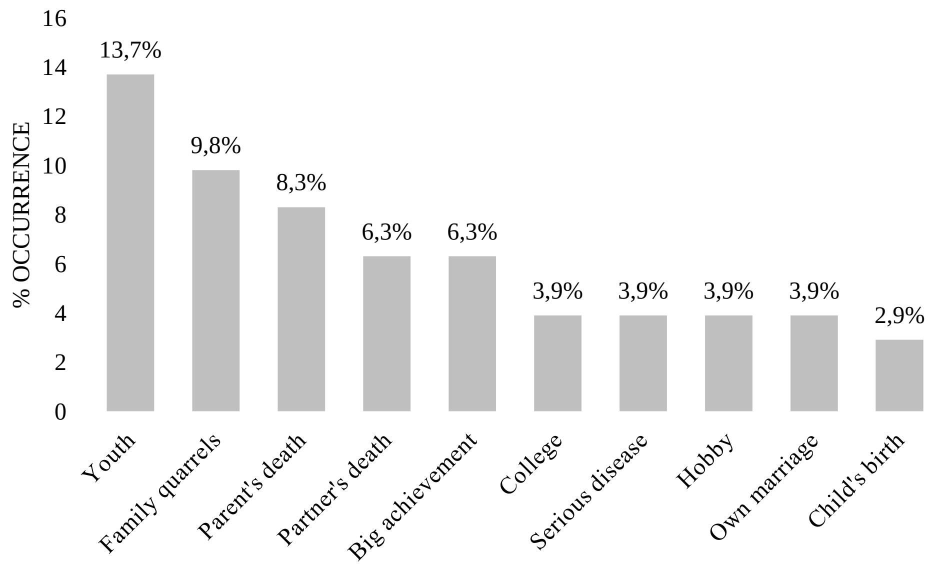


Figure 2. *Top ten occurrence of the VEM life events in memory fragments.*

3.2.3 Sentiment analysis

Lexical features of the transcripts of the participants' spoken memories were analyzed on sentiment using the PATTERN library (De Smedt & Daelemans, 2012a). The script used for sentiment analysis is shown in Appendix IV. Sentiment analysis belongs to a sub field of natural language processing that aims at the determination of the polarity (or valence) of texts. Since the focus lies on the use of affective words in emotional memories, the Dutch lexicon of PATTERN was expanded with the lexicon of Moors and colleagues (2013), with the intention to include as many affective words as possible. The mean sentiment scores were calculated for each fragment and for each separate sentence within that fragment (see Appendix IV). Sentiment scores were expressed in scores between -1 (negative valence) and 1 (positive valence) and were transformed into z-scores.

As described, transcripts were excluded from all analyses if the fragment was considered to be *just* a reflective memory, thus if it was not identified as a life event and no valence score had been assigned. Transcripts were also excluded if no emotional key words were detected after the sentiment analysis. After this, 178 of the 188 transcripts remained for further analysis with a mean sentiment score of M = .07 (SD = .15).

3.2.4 Acoustic analysis

The acoustic speech features were extracted using Praat (Boersma & Weenink, 2019). Praat is an open source software for analyzing and manipulating speech samples. Various features of F_0 (Hz) and intensity (dB) were extracted, as well as variables of voice quality and tempo features. The script used for the extraction of all acoustic features is shown in Appendix V. The set of features extracted by this script is shown in Appendix VI. Although it is not expected that F_0 and tempo features show significant associations with valence, they are still analyzed to see if results match earlier findings (Juslin & Scherer, 2005; Laukka et al., 2005; Schröder et al., 2001). These and the additional acoustic features that were expected to be closely related to valence and that were minimally interrelated to each other in this study (as established by examining the correlation matrix) are shown in Table 5.

Table 5. *Speech features extracted for acoustic analysis.*

Acoustic feature	**Variable**
F_0	Mean, standard deviation, minimum
Intensity	Mean, standard deviation, minimum
Voice quality	Hammarberg Index
Tempo	Articulation rate, speech rate, mean silent duration, mean talkspurt duration, silent rate

The Hammarberg Index (dB; Hammarberg, Fritzell, Gaufin, Sundberg, & Wedin, 1980) is a voice quality feature and gives an indication of the energy distribution in the spectrum. It is defined as the difference between the maximum energy in the lower frequency band (0 – 2000 Hz) and the maximum energy in the higher frequency band (2000 – 5000 Hz). The articulation rate is defined as the number of syllables per second, with no silences (if silences were to be included, the speech rate would have been measured). The pause rate is the number of silences per second. For each fragment, silent parts were identified by manually setting an intensity threshold in Praat (minimum silent duration = 500 milliseconds, minimum sounding duration = 150 milliseconds). The silent parts were removed in all feature extractions. Lastly, all acoustic features were normalized per speaker by transforming them into z-scores.

3.3 Procedure

Prior to the data collection, participants signed an informed consent. The AMT was introduced and to practice the retrieval of emotional memories, participants were asked to recall a specific memory for two neutral cue words (*grass* and *bread*), which were presented on a tablet. Then, participants were asked to retrieve three specific memories when presented with the cue words *sad* and *happy*. These two extreme and opposite emotions were selected to control as well as possible for the emotion elicited in the participant (Tahon et al., 2012). After the participants expressed their memory, they were asked to gather a picture or document that reminded them of that particular memory. This picture or document was then used in a follow-up meeting, which was part of the larger project. After the participants completed the AMT, they were asked to take part in the Montreal Cognitive Assessment Test (MoCA; Nasreddine et al., 2005) to check for possible mild cognitive impairments. The interviews of the 12 participants used for this study lasted between 14 and 46 minutes (M = 34m28s, SD = 10m20s). After participation, all participants received a small gift for their cooperation.

3.4 Data analysis

3.4.1 Research design

This study was an observational, cross-sectional study that analyzed data of one specific point in time, namely the moment at which the AMT was carried out. The data extracted from the AMT consisted of lexical and acoustic speech features. The lexical features consisted of the results of the sentiment analysis using PATTERN and the manual valence annotations by using the VEM. The acoustic features consisted of the results of the prosodic analysis using Praat. Both results of sentiment and acoustic analyses were first compared with the valence of the VEM life events to see if and how the lexical and acoustic speech features could predict the VEM valence outcomes. Then, the results of the sentiment analysis were compared with the results of the acoustic analysis to see how they were related to each other.

3.4.2 Statistical analysis

In order to examine how lexical and acoustic speech features relate to the valence of the described life events, a multiple regression analysis was conducted using *SPSS*. Despite the absence of normally distributed data, a parametric regression model was generated. In this model, the lexical and all acoustic features served as predictors (independent variables) for the VEM valence scores (dependent variable).

To determine the relation between *what* people say (word usage/lexical features) and *how* they say it (acoustic features), a multivariate analysis of variance (MANOVA) was conducted as well, with the results of the sentiment analysis as independent variable and the acoustic features that resulted from the regression model as dependent variables.

4. Results

In this chapter, the results of the current study will be described. First, the distribution of the data is examined. Then, the results of the multiple regression analysis are outlined. These results will serve as the basis for the subsequent analysis. In this multivariate analysis of variance, the VEM valence scores will not be taken in consideration; it purely examines the relationship between the lexical and acoustic speech features.

4.1 The prediction of valence

4.1.1 Distribution of the data

The relation between lexical and acoustic speech features and the valence of the VEM life events was examined by conducting a backwards multiple regression analysis. In this analysis, the lexical and all acoustic features served as predictors of the VEM valence scores. A backwards regression analysis was chosen, because this way the contribution of each predictor could be assessed by looking at the significance value of the *t*-test for each predictor. If a predictor did not make a statistically significant contribution to how well the model predicted the VEM valence scores, it was removed from the model and the model was re-estimated for the remaining predictors (Field, 2013a). Then, the contribution of the remaining predictors was reassessed, and so on. Ultimately, the best fitting model with the best predictors remained for further analysis. In total, 13 variables served as predictors for the VEM valence scores, 12 of which acoustic speech features (see Table 5). The remaining variable was the mean sentiment score.

First of all, the assumption of normality was tested for both the lexical and acoustic speech features. The lexical features were determined with sentiment analysis. The Kolmogorov-Smirnov test of normality showed that the sentiment scores ($D(178) = .038$, $p > .200$) did not deviate significantly from normal, which means that the data was normally distributed. The acoustic features were established using a prosodic analysis in Praat. Close to all acoustic features deviated significantly from normal, as calculated by the Kolmogorov-Smirnov test of normality. For example, the mean pitch scores significantly differed from normal distribution ($D(178) = .091$, $p = .001$). The five acoustic features that did not significantly deviate from normal, were mean intensity ($D(178) = .060$, $p > .200$), minimum intensity ($D(178) = .055$, $p > .200$), the Hammarberg Index ($D(178) = .047$, $p > .200$), articulation rate ($D(178) = .052$, $p > .200$), and silent rate ($D(178) = .066$, $p = .057$). Even though not all data was normally distributed, the parametric backwards multiple regression analysis was still conducted with all acoustic features[3].

4.1.2 Lexical and acoustic speech features as valence predictors

After the conduction of the backwards multiple regression analysis, the following outcomes became evident. The regression analysis generated a model with the best predictors after removal of seven acoustic features that did not significantly contribute to the prediction of the VEM valence scores. The final model consisted of the six valence predictors depicted in Table 6. There was no multicollinearity ($r > .9$) between predictors in this model. The model included as few predictors as possible to account for as much variation as possible (33.5%) in the VEM valence scores, without significantly deviating from models that included more predictors ($R^2 = .335$, $p = .536$). An analysis of variance (ANOVA) showed that the model was significantly better at predicting the VEM valence scores than variables that did not fit the model ($F(6, 171) = 15.876$, $p < .001$).

[3] Despite the not normally distributed data, the regression analysis was still conducted for practical reasons.

Table 6. *Best contributing variables to the prediction model of valence.*

Predictor	*β*	SE	95% CI	*p*
Mean sentiment	.496	.127	.760 – 1.263	<.001
Mean intensity	.731	.473	.535 – 2.403	.002
Intensity SD	-.524	.520	-2.137 – -.083	.034
Minimum intensity	-.480	.486	-1.958 – -.041	.041
Hammarberg Index	-.261	.159	-.851 – -.224	.001
Silent rate	.149	.139	.050 – .600	.021

The results in Table 6 show that there is a positive relationship between three predictors (mean sentiment, mean intensity, and silent rate) and the VEM valence scores. This means that if these predictors increase, the valence scores increase as well. The other predictors show a negative relationship with the VEM valence scores. This means, for instance, the lower the minimum intensity, the higher the valence scores. Furthermore, Table 6 depicts that all six predictors make a significant contribution to the model. Therefore, all predictors of this model will be used for further analysis through a multivariate analysis of variance (MANOVA) for the examination of the relation between lexical and acoustic speech features.

4.2 Relation between lexical and acoustic speech features

4.2.1 Homogeneity of covariance matrices

The previous analysis resolved which variables were applicable of predicting the VEM valence scores. In line with these results, a MANOVA was conducted to determine how lexical and acoustic speech features are related. This way, the relation could be established between *what* people say and *how* they say it. The scores of the sentiment analysis acted as the independent variable and the acoustic features shown in Table 6 served as dependent variables. The scores of the sentiment analysis were divided into two groups: positive sentiment/valence (>0, n = 123) and negative sentiment/valence (<0, n = 55). Although the sample sizes were not equal, no median split was applied to the sentiment scores, because the analysis by PATTERN (De Smedt & Daelemans, 2012a) made a dichotomous distinction between positive and negative sentiment.

First, the assumption of homogeneity of covariance matrices was tested. Box’s test of equality of covariance matrices showed that the assumption was violated because of a weak significance value (p = .048), which means that the population variance-covariance matrices of the sentiment groups (positive/negative valence) are not equal. Levene’s test of equality of error variances showed that the error variance of the mean intensity variable was not equal across the sentiment groups (p = .007), whereas the error variance of the additional dependent variables was equal across groups. Although the assumption is not met, the MANOVA was still conducted, for there is no non-parametric version of this analysis (Field, 2013b).

4.2.2 Sentiment effects on acoustic speech features

The appliance of the multivariate analysis of variance resulted in the following outcomes. The multivariate test of Pillai’s trace showed that there was no significant difference between sentiment groups (V = .057, $F(5, 172)$ = 2.094, p = .068), meaning that the positivity/negativity of the sentiment analysis had no significant effect on the acoustic speech features. Although this analysis showed no significant effects, separate univariate ANOVAs on the outcome variables revealed a significant sentiment effect on mean intensity ($F(1, 176)$ = 7.473, p = .007). Any effects on the other acoustic features, however, remained non-significant.

A second MANOVA was conducted, since the assumption of homogeneity of covariance matrices was not met due to one dependent variable only. Levene’s test had shown that the error variance of mean intensity was not equal across the sentiment groups (p = .007).

For this reason, the mean intensity variable was not included in the second MANOVA. The outcomes of this analysis were as follows. Box's test of equality of covariance matrices showed that the assumption was, in fact, met this time (p = .334), which means that the population variance-covariance matrices of the sentiment groups were equal. Levene's test of equality of error variances showed that the error variance of all acoustic features was equal across the sentiment groups (p > .050), which is in agreement with the results of Box's test of equality of covariance matrices.

Next, the multivariate test of Pillai's trace showed a significant difference between sentiment groups (V = .057, F(4, 173) = 2.632, p = .036), meaning that the positivity/negativity of the sentiment analysis had a significant effect on the remaining acoustic features. This means that if mean intensity is not included in the analysis, the combination of intensity SD (standard deviation), minimum intensity, the Hammarberg Index, and silent rate significantly relate to the sentiment score. Separate univariate ANOVAs revealed no significant sentiment effects on the acoustic features (respectively F(1, 176) = 1.217, p = .266; F(1, 176) = 2.950, p = .088; F(1, 176) = .935, p = .343; F(1, 176) = .628, p = .409). The dissimilarities between the first and second MANOVA findings are depicted in Table 7.

Table 7. *MANOVA outcomes with (1^{st}) and without (2^{nd}) mean intensity variable.*

	Box's test	**Multivariate test of Pillai's trace**				
	p	*V*	*F*	*df1*	*df2*	*p*
1^{st} MANOVA	**.048**	.057	2.094	5	172	.068
2^{nd} MANOVA	.334	.057	2.632	4	173	**.036**

Note. Significance is indicated in bold face.

5. Discussion

In this chapter, the results of the former chapter will be discussed and interpreted. The findings will be linked to the research questions and hypotheses outlined in section 2.5. Lastly, the challenges and shortcomings of the research will be examined.

5.1 Annotating valence according to the VEM

Since emotions and memories are highly subjective and personal, earlier studies seemed to struggle with obtaining valence information using objective measures that do not take personal interpretations into account (e.g., De Smedt & Daelemans, 2012b; Moors et al., 2013; Scherer & Oshinsky, 1977; Schröder et al., 2001; Tahon et al., 2012). The first sub question (*How can valence be annotated in a reliable manner in spontaneous spoken emotional memories?*) of the current study related to this problem. A novel valence coding scheme, the Valence of Emotional Memories scale (VEM), was developed to annotate the valence of emotional autobiographical memories. The VEM determines the valence of a certain memory while allowing subjective interpretations (by including a subjectivity score), which is a new development in emotional research. In contrast to studies that focused on acoustic speech features, such as pitch, intensity, and speech rate to objectively measure the valence of affective speech (Busso & Rahman, 2012; Goudbeek & Scherer, 2010; Laukka et al., 2005; Schröder et al., 2001; Tahon et al., 2012), the VEM requires human interpretation and describes valence according to the content of speech and life events that are described by participants. The advantage of the VEM is that it respects the subjective and personal emotional memories of the participants, while being flexible in adding life events to annotate memories that may not be represented in the VEM as depicted in Table 4. A moderate inter-rater agreement showed that the VEM is a decent measure for annotating valence in spontaneous spoken emotional memories.

Some remarks should be made about the newly developed VEM. One disadvantage of the VEM is that it *only* takes word usage into account. For optimal valence detection, all sorts of variables should be taken into consideration; not only lexical and acoustic features, but, for instance, nonverbal features, facial expressions, and body postures (Jacob-Dazarola et al., 2016) as well. More research should be carried out to determine the value of such amplifications. Furthermore, since the VEM is a novel scheme for annotating valence, it cannot be considered a standardized test for detecting valence (yet). Additional research should be conducted to ensure the value of the VEM as a valid emotional valence measurement. It should also be noted that since the VEM is the first valence coding scheme that takes the subjectivity of personal memories into account, future research should reveal if the subjectivity score is a sufficient approach to orderly manipulate the initial VEM valence scores as depicted in Table 4. The two independent raters of the emotional fragments occasionally found themselves in the position of doubting the accuracy of the subjectivity score, so perhaps a score of +1/-1 should be re-evaluated in additional research in a way that the subjectivity score can be applied to any situation and any life event more precisely.

In addition, some difficulties arose while annotating the valence of the emotional fragments. When a participant described his/her memory, it was sometimes challenging to assign the right life event to a fragment. For instance, if a participant mentioned he/she went on a spectacular hiking trip for the first time in life, is the most appropriate life event *travelling* (since it took place on vacation), *hobby* (hiking is assumed to be a hobby), or *big achievement* (due to the personal victory)? Issues like these were likely to lower the inter-rater agreement between the two independent raters. Thus, the subjective judgment of the raters has both an upside and a downside: there is room for manipulation of the valence score (using the subjectivity score), but the different interpretations of the life events may be the cause of a lower agreement between raters.

5.2 Predicting valence through lexical and acoustic speech features

The relation between lexical and acoustic speech features and the valence of life events as described by the VEM was examined to identify which features succeeded at predicting valence. This way, it could be established if and how the valence of lexical and acoustic features related to the VEM valence scores. A backwards multiple regression analysis generated a model that included the sentiment scores and five acoustic features for significantly predicting the VEM valence scores. As expected, four of these acoustic features were parameters of intensity (mean, standard deviation, and minimum) and voice quality (Hammarberg Index). This is in line with results from Schröder and colleagues (2001) and Laukka and colleagues (2005), who showed valence can be associated with intensity, and Tahon and colleagues (2012), who stated that voice quality features were more useful for valence detection than additional acoustic features. The one acoustic feature that was not expected to be associated with valence, but did significantly predict the VEM valence scores, was silent rate. The earlier mentioned studies did find some tempo features to be related to valence (Laukka et al., 2005; Schröder et al., 2001), but none of them specifically mentioned silent rate as a predicting variable for valence. This could be explained by a difference in methodology (they might, for example, not have made the distinction between silent and sounding parts in their audio samples, while this was the case in the current study), but such dissimilarities are not specified by the authors.

The model generated by the regression analysis contained the most fitting predictors after the removal of seven variables that did not significantly contribute to the prediction of the VEM valence scores. As expected, most of the excluded variables involved F_0 features (mean, standard deviation, and minimum) and tempo features (articulation rate, speech rate, mean silent duration, and mean talkspurt duration). All studies described in section 2.4 reported correlations between F_0 features and valence, but these findings were very inconsistent and often contradictory. For this reason, it was expected that F_0 features are not suitable for predicting VEM valence scores and this expectation has shown to be true for the dataset of this study.

The regression analysis showed that word usage (lexical features) was positively and significantly associated with the VEM valence scores. This means that the PATTERN sentiment analysis (De Smedt & Daelemans, 2012b), that was used to extract the lexical features, is a significant predictor of valence. The more positive the sentiment score of a given fragment, the more positive the VEM valence score of that same fragment. Thus, the more (combinations of) positive key words a fragment contained, the more positive was the life event described by the participant. This outcome is as expected, since the VEM valence scores and the sentiment analysis were both based on the content of speech, in contrast to the prosodic analysis that was based on acoustic speech features. This means that the expectation that the outcomes of the sentiment analysis are positively related to the VEM valence scores is true for this dataset.

Of the five significant acoustic features that were included in the regression model, a positive relation occurred between two features and the VEM valence scores, among which mean intensity. This means that if the mean intensity increased, the VEM valence scores increased as well. Thus, if in a particular fragment the participant spoke louder, the life event described by the participant was more positive. This outcome is not in agreement with earlier findings. Goudbeek and Scherer (2010) found that positive valence could be linked to lower levels of intensity. On the other hand, Schröder and colleagues (2001) did report increased intensity levels, but in relation to *negative* valence values. There is no straightforward explanation for the latter dissimilarity, but it should be mentioned that the correlations of Schröder and colleagues (2001) were “less numerous and less strong”, in contrast to the associations of the current study. The discrepancy between the results of Goudbeek and Scherer (2010) and the current study could be explained by the fact that the former controlled for level of arousal. They found that positive valence was only associated with a lower level of intensity at high levels of arousal. The current study, however, did not control for arousal. This raises the question if the same results would have been found if arousal had been controlled for in this study. This is a variable that should be taken into account in future research.

Silent rate showed a positive relation to the VEM valence scores as well. This means that the more silences or pauses (of > .5 seconds) appeared in the fragment, the more positive was the life event described by the participant. Initially, this seems like an illogical outcome, but Schröder and colleagues (2001) showed that *negative* valence was associated with *longer*, fewer pauses. A positive silent rate, on the other hand, gives the indication that a fragment contains *more*, but shorter silences. This relation between silences and valence has not been shown in affective speech research before, but can be rationalized in the light of Schröder and colleagues' (2001) results.

The three remaining acoustic variables that were significantly associated with the VEM valence scores (intensity SD, minimum intensity, and the Hammarberg Index) showed negative relationships. For the intensity SD, this means that if the intensity SD decreased, the VEM valence score increased. Thus, if the variation in intensity lessened, the life event described by the participant was more positive. This result is in line with findings of Goudbeek and Scherer (2010), who stated that positive valence was associated with less intensity variation. However, when they controlled for level of arousal, Goudbeek and Scherer (2010) found deviating results. At low levels of arousal, positive valence was associated with more variation in intensity and at high levels of arousal, they associated *negative* valence with more intensity variation. It would be interesting to see if comparable findings arise when arousal is controlled for in future research.

Minimum intensity was also negatively related to the VEM valence scores, meaning that if the minimum intensity decreased, the VEM valence score increased. Hence, the weaker the voice was at its weakest point in the fragment, the more positive was the life event described by the participant. Although this outcome has not been shown in the studies mentioned before, Schröder and colleagues (2001) did report more prominent intensity *maxima* in association with *negative* valence.

The fifth and last acoustic feature that was significantly associated with the VEM valence scores was the Hammarberg Index, a voice quality feature that describes the difference between the maximum energy in the lower and higher frequency band (Hammarberg et al., 1980). The Hammarberg Index was negatively related to the VEM valence scores, meaning that if the spectral slope decreased (was less steep), the VEM valence score increased. Thus, the more energy was located in the higher frequency band (and thus the higher the vocal effort), the more positive the life event described by the participant. This outcome is not comparable with the findings of Goudbeek and Scherer (2010), who found that the Hammarberg Index correlated positively with valence (meaning a steeper spectral slope was associated with positive valence). A possible explanation could be that the data used in this study was gathered by eliciting happy and sad memories that did not just differ in valence, but in arousal as well. Since the current study did not control for level of arousal (or any other dimension), it is unknown to which degree the valence scores are confounded with such dimensions. Schröder and colleagues (2001) showed that a flatter spectral slope can indeed be associated with high arousal levels. Therefore, the result of the current study of a lower Hammarberg Index associating with positive valence could actually be a result that is related to arousal rather than valence.

In general, the model generated by the regression analysis included six predictors of the VEM valence scores: mean sentiment score, mean intensity, intensity SD, minimum intensity, the Hammarberg Index, and silent rate. These predictors together accounted for 33.5% of the variation in the VEM valence scores, meaning that 66.5% of the variation remains unidentified.

This could be explained by independent factors that were not taken into consideration in the current study, but could have had influence on the outcome variables, such as the previously mentioned level of arousal. An additional factor that could have influenced the variation in the VEM valence scores, but was not taken into consideration in this study, is individual differences. A different statistical analysis, such as a linear mixed-effects regression analysis, could be used in future research to control for acoustic variation due to the individual participant.

Overall, the expectations of the relation between lexical and acoustic speech features and the VEM valence scores can be confirmed. As expected, the results showed a positive, significant relation between the lexical features and the VEM valence scores, and acoustic features such as voice quality and intensity were also significantly associated with the VEM valence scores. In relation to these findings, it should be noted, however, that not all data used in the regression analysis was normally distributed. This means that an important assumption of the statistical analysis has been violated and that the results of this analysis should be interpreted with caution. To prevent this in the future, a bigger data sample could be used or an alternative, non-parametric statistical analysis could be conducted.

5.3 Lexical and acoustic features in affective speech

The relation between the valence of lexical and acoustic speech features was investigated to establish if *the way* something is said (acoustic features) compares to *what* has been said (lexical features). To examine this, all six significant valence predictors of the regression model were used for further analysis through a MANOVA. Two MANOVAs were conducted. The first one included all five acoustic features as the dependent variables and the word usage (the outcomes of the sentiment analysis) as the independent variable. No significant effects were found after this analysis. A second MANOVA was conducted without mean intensity as one of the dependent variables. This time, the analysis showed a significant difference between sentiment groups (positive/negative valence). This means that if mean intensity is not included in the analysis, the combination of intensity SD, minimum intensity, the Hammarberg Index, and silent rate were significantly associated with the sentiment score. This result is in line with the expectation that the valence of lexical and acoustic speech features match, since the acoustic features can be influenced by sentiment scores. However, the nature of this influence is unclear, since the separate ANOVAs showed no significant outcomes. Furthermore, only four of the original 13 acoustic features (see Table 5) were taken into account in the second MANOVA, since those variables were significantly associated with the VEM valence scores. For this reason, it can only be concluded that the intensity SD and minimum intensity of the voice, the Hammarberg Index, and silent rate (*how* something has been said) significantly depend on the valence of *what* has been said (word usage).

5.4 Limitations of the current study

A few shortcomings of the current study have been mentioned previously. First of all, the current study did not control for levels of arousal, in contrast to Goudbeek and Scherer (2010). Due to this, it is unknown to which degree the VEM valence scores can be attributed to the actual valence dimension. This drawback may be an explanation for various incomparable outcomes of the current study, such as the negative relation between the Hammarberg Index and the VEM valence scores. Besides this, the study did not control for the individual participant, which may have caused the relatively low variation accounted for (33.5%) in the model of the VEM valence predictors that was generated by the regression analysis. A third limitation could be that audio fragments were excluded from sentiment and acoustic analyses if the VEM valence scores were considered to be neutral (valence scores between 3-5). Only fragments with negative (1-3) and positive (5-7) valence scores were included, because this study made the first step to the comprehension of valence in spontaneous affective speech production of older adults. Additional research that incorporates all data (including valence scores of a neutral nature) should reveal a more thorough understanding of this research field.

Furthermore, the data of this study was collected using the AMT, which provided spontaneous speech samples of emotional memories. The advantages of spontaneous speech samples are the increased likelihood of obtaining ecologically valid speech samples and the preservation of the natural context of vocal expression (Juslin & Scherer, 2005). There are, however, disadvantages of working with natural speech, as described in section 2.1.3. These obstacles, such as the condition of the audio samples (since they were recorded at a location of the participant's choice), might have played their part in the current study, too. It is also difficult to determine exactly which emotion was felt by the participants. In the course of the AMT, participants were asked to briefly describe sad and happy memories, but people can feel differently towards different situations and may have a dissimilar interpretation of sadness and happiness. Nonetheless, the elicitation of emotions was managed and regulated as much as possible, and the use of spontaneous speech samples is still preferred over acted or portrayed emotional speech samples.

5.5 Suggestions for future research

In the light of the current study, some questions remain unanswered. Although the study did aim at the inclusion of more acoustic features than just common voice cues such as pitch, intensity, and speech rate to obtain valence differentiation, only one additional acoustic feature for voice quality (the Hammarberg Index) was incorporated in the analyses. For future research, it would be interesting to include more voice quality features, such as harmonics-to-noise ratio, jitter, and shimmer. Tahon and colleagues (2012) also included two novel voice quality features in their study: the relaxation coefficient, that indicated the relaxation of the voice, and the functions of phase-distortion, that indicated the distortion of the phase spectrum around its linear component. As described in section 2.4, they concluded that these features could be useful for emotional valence detection, too. Additionally, Laukka and colleagues (2005) suggested that valence could be associated with acoustic features like speech rhythm, since they found that the vocal cues associated with valence were more independent of autonomic physiological changes, and that valence perception may have become better differentiated as the length of the spoken phrase increased. Since different acoustic features seem to be associated with both arousal and valence, future studies should try to find other, new cues for detecting valence instead of focusing on acoustic features that we now know are more closely related to arousal than to valence.

Furthermore, the current study contributed to older adult affective speech research, but there are more questions in this field to be answered. Since the confirmed presence of the positivity effect in older adults (Birditt & Fingerman, 2003; Kennedy et al., 2004; Pasupathi et al., 1998; Reed et al., 2014; Schieman, 1999), it would be valuable to search for the appearance of this effect in lexical and acoustic speech features. Results of younger and older adults could be compared to discover if the valence of lexical and acoustic features is in fact higher in older adult speech. These findings could lead to a broader understanding of the role of valence in spontaneous affective speech production of older adults.

6. Conclusion

This study explored the field of spontaneous affective speech production by focusing on the valence of lexical and acoustic speech features. To broaden this research field, the spontaneous speech production of older adults was examined using the VEM, a novel valence scheme for annotating emotional autobiographical memories. The main research question of the study was as follows:

> *How can valence be measured in spontaneous emotional speech production of healthy older adults (≥ 65 years of age)?*

To answer this research question, various sub questions were constructed. After examination of the first sub question (*How can valence be annotated in a reliable manner in spontaneous spoken emotional memories?*), it became evident that the VEM is a decent measure for annotating valence in spontaneous spoken emotional memories. It should, however, be noted that the use of the VEM is still in its infancy and that additional research should be conducted to ensure the value of the VEM as a valid emotional valence measurement.

The second sub question was as follows: *To what extent can sentiment analysis predict valence in spontaneous spoken emotional memories?* Results showed that word usage is an adequate predictor for the valence of speech production. This outcome was in line with the expectation that the results of the sentiment analysis were positively related to the valence scores obtained by the VEM. Furthermore, in relation to the third sub question (*To what extent can acoustic speech features predict valence in spontaneous spoken emotional memories?*), it was expected that the VEM valence scores could be associated with particular acoustic features, namely features of intensity and voice quality. This expectation also appeared to be true; three intensity features, one voice quality feature, and one tempo feature significantly predicted the VEM valence scores.

Lastly, the relation between the valence of lexical and acoustic features was investigated to compare *how* something is said to *what* has been said. To this extent, the fourth and last sub question was as follows: *How are automated analyses of lexical and acoustic features in speech production related to each other with respect to valence expression in spontaneous spoken emotional memories?* Expectations were that a significant association between the word usage and acoustic speech features would be present. This expectation was met, but only for four acoustic features. Eventually, it was established that the intensity SD and minimum intensity of the voice, the Hammarberg Index, and silent rate (*how* something had been said) significantly depended on the valence of *what* had been said (word usage). No indisputable statements could be made about the association between lexical and *all* acoustic features in affective speech.

Overall, it seems that the valence of spontaneous emotional speech production of healthy older adults (≥ 65 years of age) can successfully be measured according to a novel valence annotation scheme. Lexical and acoustic features seem to be useful for detecting valence in affective speech. However, additional research should be conducted to determine a more detailed relation between *what* is said and *how* it has been said. In conclusion, it can be stated that the outcomes of this study contribute to the already ambiguous field of valence detection in spontaneous speech production. Some results did and some results did not compare to the (inconsistent) findings of previous studies. This study initiated the step towards a thorough understanding of valence in spontaneous affective speech production of older adults. Subsequent time will unveil how emotional dimensions, in particular valence, can best be described in older adult speech.

References

Andrade, E.B., & Cohen, J.B. (2007). On the consumption of negative feelings. *Journal of Consumer Research, 34*, 283-300.

Aubergé, V., Audibert, N., & Rilliard, A. (2006). Auto-annotation: An alternative method to label expressive corpora. In *Proceedings of LREC.*

Bachorowski, J-A. (1999). Vocal expression and perception of emotion. *Current Directions in Psychological Science, 8*, 53-57.

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

Barrett, L.F. (2006a). Emotions as natural kinds? *Perspectives on Psychological Science, 1*, 28-58.

Barrett, L.F. (2006b). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review, 10*(1), 20-46.

Barrett, L.F. (2011a). Was Darwin wrong about emotional expressions? *Current Directions in Psychological Science, 20*, 400-406.

Barrett, L.F. (2011b). Bridging toke identity theory and supervenience theory through psychological construction. *Psychological Inquiry, 22*, 115-127.

Barrett, L.F. (2016). Navigating the science of emotion. In H. Meiselman (Ed.), *Emotion Measurement* (pp. 31-63). Oxford: Elsevier.

Berntsen, D., & Rubin, D.C. (2002). Emotionally charged autobiographical memories across the life span: The recall of happy, sad, traumatic, and involuntary memories. *Psychology and Aging, 17,* 636-652.

Berntsen, D., & Rubin, D.C. (2004). Cultural life scripts structure recall from autobiographical memory. *Memory and Cognition, 32*(3), 427-442.

Birditt, K., & Fingerman, K.L. (2003). Age and gender differences in adults' descriptions of emotional reactions to interpersonal problems. *Journal of Gerontology Series B: Psychological Sciences, 58B,* 237-245.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer [Computer program]. Version 6.0.52, retrieved May 2nd 2019 from http://www.praat.org/.

Breitenstein, C., Van Lancker, D., & Daum, I. (2001). The contribution of speech rate and pitch variation to the perception of vocal emotions in a German and an American sample. *Cognition and Emotion, 15*, 57-79.

Busso, C., & Rahman, T. (2012). Unveiling the acoustic properties that describe the valence dimension. In *Proceedings of Interspeech 2012* (pp. 1179-1182). Portland, USA.

Conway, M.A. (2005). Memory and the self. *Journal of Memory and Language, 53*, 594-628.

De Smedt, T., & Daelemans, W. (2012a). Pattern for Python. *Journal of Machine Learning Research, 13*, 2063-2067.

De Smedt, T., & Daelemans, W. (2012b). "Vreselijk mooi!" (terribly beautiful): A subjectivity lexicon for Dutch adjectives. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)* (pp. 3568-3572). Istanbul, Turkey.

Douglas-Cowie, E., Campbell, N., Cowie, R., & Roach, P. (2003). Emotional speech: Towards a new generation of data bases. *Speech Communication, 40,* 33-60.

Eichhorn, J.T., Kent, R.D., Austin, D., & Vorperian, H.K. (2017). Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice, 32*(5), 644.e1-644.e9.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169-200.

Elfenbein, H.A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin, 128*(2), 203-235.

Ellsworth, P.C., & Scherer, K.R. (2003). Appraisal processes in emotion. In R.J. Davidson, H.H. Goldsmith, & K.R. Scherer (Eds.), *Handbook of Affective Sciences* (pp. 572-595). New York: Oxford University Press.

Erdoğan, A., Baran, B., Avlar, B., Taş, A.Ç., & Tekcan, A.I. (2008). On the persistence of positive events in lifescripts. *Applied Cognitive Psychology, 22*, 95-111.

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

Feldman, L.A. (1993). Distinguishing depression from anxiety in self-report: Evidence from confirmatory factor analysis on nonclinical and clinical samples. *Journal of Consulting and Clinical Psychology, 61,* 631-638.

Field, A. (2013a). Regression. In A. Field (Ed.), *Discovering Statistics using IBM SPSS Statistics* (4th ed., pp. 293-356). Los Angeles: Sage.

Field, A. (2013b). Multivariate analysis of variance (MANOVA). In A. Field (Ed.), *Discovering Statistics using IBM SPSS Statistics* (4th ed., pp. 623-664). Los Angeles: Sage.

Fontaine, J.R.J., Scherer, K.R., Roesch, E.B., & Ellsworth, P.C. (2007). The world of emotions is not two-dimensional. *Association for Psychological Science, 18*(12), 1050-1057.

Frijda, N.H. (1986). *The emotions.* Cambridge: Cambridge University Press.

Gobl, C., & Chasaide, A.N. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication, 40*(1-2), 189-212.

Goudbeek, M., & Scherer, K.R. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America, 128,* 1322-1336.

Glück, J., & Bluck, S. (2007). Looking back across the life span: A life story account of the reminiscence bump. *Memory and Cognition, 35,* 1928-1939.

Grysman, A., & Dimakis, S. (2017). Later adults' cultural life scripts of middle and later adulthood. *Aging, Neuropsychology, and Cognition, 25*(3), 406-426.

Habermas, T., & Köber, C. (2015). Autobiographical reasoning is constitutive for narrative identity: The role of the life story for personal continuity. In K.C. McLean, & M. Syed (Eds.), *The Oxford Handbook of Identity Development* (pp. 149-165). New York: Oxford University Press.

Hammarberg, B., Fritzell, B., Gaufin, J., Sunberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica, 90*(1-6), 441-451.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34,* 115-139.

Hunter, P.G., Schellenberg, E.G., & Schimmack, U. (2008). Mixed affective responses to music with conflicting cues. *Cognition and Emotion, 22*, 327-352.

Izard, C.E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science, 2,* 260-280.

Jacob-Dazarola, R., Ortíz-Nicolás, J.C., & Cárdenas, L. (2016). Behavioral measures of emotion. In H. Meiselman (Ed.), *Emotion Measurement* (pp. 101-124). Amsterdam: Woodhead Publishing.

James, W. (1884). What is an emotion? *Mind, 9*, 188-205.

Janssen, S.M.J. (2015). Is there a cultural life script for public events? *Applied Cognitive Psychology, 29,* 61-68.

Janssen, S.M.J., Murre, J.M.J., & Meeter, M. (2008). Reminiscence bump in memory for public events. *European Journal of Cognitive Psychology, 20*, 738-764.

Janssen, S.M.J., & Rubin, D.C. (2011). Age effects in cultural life scripts. *Applied Cognitive Psychology, 25,* 291-298.

Johnstone, T., & Scherer, K.R. (2000). Vocal communication of emotion. In M. Lewis, & J.M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 220-235). New York: Guilford Press.
Juslin, P.N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770-814.
Juslin, P.N., & Scherer, K.R. (2005). Vocal expression of affect. In J.A. Harrigan, R. Rosenthal, & K.R. Scherer (Eds.), *The New Handbook of Methods in Nonverbal Behavior Research,* pp. 65-135. Oxford: Oxford University Press.
Kanluan, I., Grimm, M., & Kroschel, K. (2008). Audio-visual emotion recognition using an emotion recognition space concept. In *Proceedings of the 16th European Signal Processing Conference.*
Kennedy, Q., Mather, M., & Carstensen, L.L. (2004). The role of motivation in the age-related positivity effect in autobiographical memory. *American Psychological Society, 15*(3), 208-214.
Labouvie-Vief, G., Grühn, D., & Studer, J. (2010). Dynamic integration of emotion and cognition: Equilibrium regulation in development and aging. In R.M. Lerner, M.E. Lamb, & A.M. Freund (Eds.), *The Handbook of Life-span Development* (Vol. 2, pp. 79-115). Hoboken: Wiley.
Larsen, J.T., McGraw, A.P., & Cacioppo, J.T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81,* 684-696.
Laukka, P., Juslin, P.N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion, 19*(5), 633-653.
Leinonen, L., Hiltunen, T., Linnankoski, I., & Laakso, M-L. (1997). Expression of emotional-motivational connotations with a one-word utterance. *Journal of the Acoustical Society of America, 102*(3), 1853-1863.
Linville, S.E. (2001). *Vocal ageing*. San Diego: Singular Publishing Group.
Luchetti, M., & Sutin, A.R. (2018). Age differences in autobiographical memory across the adult lifespan: Older adults report stronger phenomenology. *Memory, 26*(1), 117-130.
Magai, C., Consedine, N.S., Krivoshekova, Y.S., Kudadjie-Gyamfi, E., & McPherson, R. (2006). Emotion experience and expression across the adult life span: Insights from a multimodal assessment study. *Psychology and Aging, 21*(2), 303-317.
Matt, G.E., Vasquez, C., & Campbell, W.K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12,* 227-255.
Mauss, I.B., & Robinson, M.D. (2009). Measures of emotion: A review. *Cognition and Emotion, 23*(2), 209-237.
McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282.
Meeter, M., & Hendriks, M. (2012). Geheugen. In R. Kessels, P. Eling, R. Ponds, J. Spikman, & M. Van Zandvoort (Eds.), *Klinische neuropsychologie* (pp. 197-218). Amsterdam: Boom.
Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A-L., De Schryver, M., De Winne, J., & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods, 45,* 169-177.
Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society, 53*(4), 695-699.
Ortony, A., & Turner, T.J. (1990). What's basic about basic emotions? *Psychological Review, 97*(3), 315-331.

Pasupathi, M., Carstensen, L.L., Turk-Charles, S., & Tsai, J. (1998). Emotion and aging. *Encyclopedia of Mental Health, 2*, 91-101.

Pereira, C. (2000). Dimensions of emotional meaning in speech. In R. Cowie, E. Douglas-Cowie, & M. Schröder (Eds.), *Proceedings of the ISCA Workshop on Speech and Emotion* (CD-ROM). Belfast: International Speech Communication Association.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmr, G., & Vesely, K. (2011). The Kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding,* IEEE Signal Processing Society.

Protopapas, A., & Lieberman, P. (1997). Fundamental frequency of phonation and perceived emotional stress. *Journal of the Acoustical Society of America, 101*(4), 2267-2277.

Raes, F., Hermans, D., Williams, J.M.G., Beyers, W., Brunfaut, E., & Eelen, P. (2006). Reduced autobiographical memory specificity and rumination in predicting the course of depression. *Journal of Abnormal Psychology, 115*(4), 699-704.

Reed, A.E., Chan, L., & Mikels, J.A. (2014). Meta-analysis of the age-related positivity effect: Age differences in preferences for positive over negative information. *Psychology and Aging, 29*(1), 1-15.

Reubold, U., Harrington, J., & Kleber, F. (2010). Vocal aging effects on F_0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication, 52*, 638-651.

Roessler, R., & Lester, J.W. (1976). Voice predicts affect during psychotherapy. *Journal of Nervous and Mental Disease, 163*, 166-176.

Rubin, D.C., & Berntsen, D. (2003). Life scripts help to maintain autobiographical memories of highly positive, but not highly negative, events. *Memory and Cognition, 31*, 1-14.

Rubin, D.C., Berntsen, D., & Hutson, M. (2009). The normative and the personal life: Individual differences in life scripts and life story events among USA and Danish undergraduates. *Memory, 17*(1), 54-68.

Russell, J.A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145-172.

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

Rusting, C.L. (1998). Personality, mood, and cognitive processing of emotional information: Three conceptual frameworks. *Psychological Bulletin, 124*, 165-196.

Schank, R.C., & Abelson, R.P. (1977). Scripts, plan, and knowledge. In P.N. Johnson-Laird, & P.C. Wason (Eds.), *Thinking, Readings in Cognitive Science* (pp. 421-435). Cambridge: Cambridge University Press.

Scherer, K.R. (1972). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal Communication* (pp. 105-111). New York: Oxford University Press.

Scherer, K.R. (1982). Emotion as a process: Function, origin, and regulation. *Social Science Information, 21*, 555-570.

Scherer, K.R. (1994). Toward a concept of 'modal emotions'. In P. Ekman & R.J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 25-31). New York: Oxford University Press.

Scherer, K.R. (2004). Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them? *Journal of New Music Research, 33*(3), 239-251.

Scherer, K.R. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695-729.

Scherer, K.R., & Oshinsky, J.S. (1977). Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion, 1*(4), 331-346.

Schieman, S. (1999). Age and anger. *Journal of Health and Social Behavior, 40*, 273-289.

Schimmack, U., & Colcombe, S. (2007). Eliciting mixed feelings with the paired-picture paradigm: A tribute to Kellogg (1915). *Cognition and Emotion, 21*, 1546-1553.

Schmidt, J., Janse, E., & Scharenborg, O. (2016). Perception of emotion in conversational speech by younger and older listeners. *Frontiers in Psychology, 7*, 781.

Schröder, M., Cowie, R., Douglas-Cowie, E., Westerdijk, M., & Gielen, S. (2001). Acoustic correlates of emotion dimensions in view of speech synthesis. In *Proceedings of Eurospeech 2001* (Vol. 1), pp. 87-90. Aalborg: International Speech Communication Association.

Shamaskin, A.M., Mikels, J.A., & Reed, A.E. (2010). Getting the message across: Age differences in the positive and negative framing of health care messages. *Psychology and Aging*, *25*, 746-751.

Shuman, V. & Scherer, K.R. (2015). Emotions, psychological structure of. In J.D. Wright (Ed.), *International Encyclopedia of the Social and Behavioral Sciences* (2nd ed., Vol. 7, pp. 526-533). Oxford: Elsevier.

Spaniol, J., Voss, A., & Grady, C.L. (2008). Aging and emotional memory: Cognitive mechanisms underlying the positivity effect. *Psychology and Aging, 23,* 859-872.

Tahon, M., Degottex, G., & Devillers, L. (2012). Usual voice quality features and glottal features for emotional valence detection. *Speech Prosody*, 693-696.

Van Bezooijen, R. (1984). *Characteristics and recognizability of vocal expressions of emotion*. Dordrecht: Foris.

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: Two experience sampling studies. *Emotion, 9*, 83-91.

Williams, J.M.G. (1996). Depression and the specificity of autobiographical memory. In D.C. Rubin (Ed.), *Cognition, Emotion and Psychopathology* (pp. 251-269). Cambridge: Cambridge University Press.

Williams, J.M.G., & Broadbent, K. (1986). Autobiographical memory in suicide attempters. *Journal of Abnormal Psychology, 95*(2), 144-149.

Xu, R., Yang, J., Feng, C., Wu, H., Huang, R., Yang, Q., Li, Z., Xu, P., Gu, R., & Luo, Y-J. (2018). Time is nothing: Emotional consistency of autobiographical memory and its neural basis. *Brain Imaging and Behavior, 12,* 1053-1066.

Yik, M.S.M., Russell, J.A., & Feldman-Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology, 77*, 600-619.

Appendices

Appendix I: Participants' demographics

Participant	**Gender**	**Age**	**Education** (Dutch)
2	Male	68	MULO/MMS/MAVO/MBO
5	Male	79	MULO/MMS/MAVO/MBO
6	Female	77	MULO/MMS/MAVO/MBO
8	Male	77	MULO/MMS/MAVO/MBO
9	Male	79	Ambachtsschool
15	Male	72	Universiteit/hoger onderwijs
17	Female	69	Universiteit/hoger onderwijs
19	Female	81	Universiteit/hoger onderwijs
20	Male	71	MULO/MMS/MAVO/MBO
21	Female	66	MULO/MMS/MAVO/MBO
22	Female	69	Universiteit/hoger onderwijs
23	Male	68	Universiteit/hoger onderwijs

Appendix II: Information letter received by participants

UNIVERSITY
OF TWENTE.

Informatie over de studie 'Emoties en Ouderen'

Geachte heer/mevrouw,

U heeft aangegeven geïnteresseerd te zijn in deelname aan de studie 'Emoties en Ouderen'. Wij willen u alvast bedanken voor uw openheid en interesse. In deze informatiebrief staat meer informatie over de studie. We leggen uitgebreid uit wat deelname voor u betekent. U kunt deze informatie nog eens rustig doorlezen en eventuele vragen die u heeft opschrijven. Bespreek het met uw familie en vrienden. Als u vragen heeft, kunt u terecht bij de onderzoekers, […] en […], of de projectleider, […]. U vindt de contactgegevens onderaan deze brief.

Wat is het doel van het project?
Het eerste doel van onze studie is te onderzoeken hoe herinneringen van *mensen van 65 jaar of ouder* positieve of negatieve emoties kunnen oproepen.

Het tweede doel van deze studie is om te onderzoeken of we emoties kunnen herkennen op basis van verschillende metingen: uw gezichtsuitdrukkingen, uw spraak en een aantal lichamelijke signalen zoals hartslag, lichamelijke activiteit (beweging) en huidgeleiding.

Wanneer kunt u deelnemen aan de studie?
Voor deze studie zoeken wij mensen die 65 jaar of ouder zijn zonder noemenswaardige geheugenproblemen. Daarnaast is het belangrijk dat u de Nederlandse taal kunnen lezen en begrijpen. Verder is het belangrijk dat u goed kunt zien en horen, eventueel met bril of gehoorapparaat. Een laatste punt is dat u geen pacemaker heeft.

Wat houdt dit voor u in?
Na het ontvangen van deze informatiebrief zullen wij u binnen een aantal dagen bellen om te vragen of u interesse heeft om mee te doen aan deze studie. Tijdens dit telefonisch gesprek kunt u al uw vragen stellen aan de onderzoekers. Natuurlijk mag u ook eerder of juist later nog een keer bellen voor meer informatie of om aan te geven dat u wilt meedoen. Als u aangeeft dat u wilt meedoen aan de studie zullen we samen een eerste afspraak inplannen. Deze eerste afspraak zal bij u thuis of in een voor u bekende omgeving plaatsvinden en is bedoeld om kennis te maken met elkaar. Deze afspraak zal ongeveer anderhalf uur duren.

Als voorbereiding voor de eerste afspraak vragen we u een aantal foto's of documenten te verzamelen waaraan u een positieve of negatieve herinnering heeft. U kunt bij documenten denken aan bijvoorbeeld geboortekaartjes, trouwuitnodigingen of krantenartikelen maar geen voorwerpen die niet in een boek kunnen worden verwerkt. Tijdens de eerste afspraak zullen we u vragen om positieve en negatieve herinneringen te benoemen en deze te laten zien door middel van een foto of document. We zullen deze digitaal opslaan voor uw levensalbum. Een levensalbum is een boek waarin mensen hun levensverhaal vastleggen. Door middel van persoonlijke foto's of documenten kunt u uw herinneringen opslaan en bekijken in het boek. U kunt bijvoorbeeld een herinnering hebben van een gebeurtenis waarin u iets positiefs overkomt of een moment in uw leven waar het u tegen zat. Denkt u bijvoorbeeld aan een foto uit uw jeugd waar u een positieve herinnering aan heeft. Met deze persoonlijke herinneringen en bijbehorende foto's maken wij een digitaal levensalbum dat u doet herinneren aan een aantal van deze gebeurtenissen.

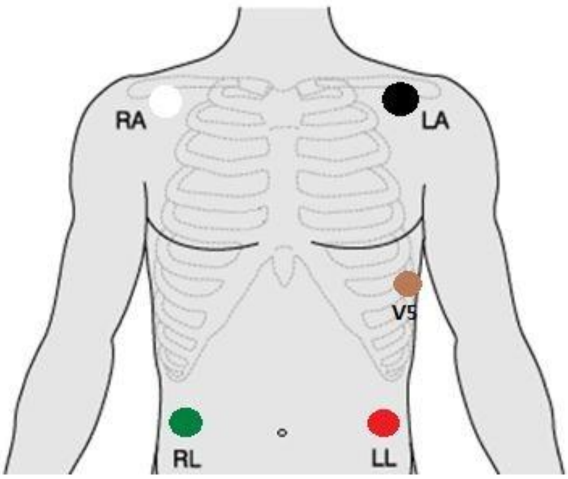
Ook zullen er tijdens deze afspraak drie vragenlijsten worden afgenomen. Tijdens deze afspraak zullen we alleen audio-opnames maken. Tot slot plannen we een tweede afspraak in. Die zal ongeveer één tot twee weken na de eerste afspraak zijn.

Tijdens de tweede afspraak willen we u gaan interviewen over uw levensalbum en de foto's, documenten en verhalen die u heeft uitgekozen. Ter vergelijking leggen we ook een aantal foto's voor die niets te maken hebben met uw herinneringen, dit zijn algemene foto's. Deze afspraak zal maximaal anderhalf uur duren. Tijdens deze afspraak zullen we audio- en video-opnames maken en metingen uitvoeren van de hartslag, beweging en huidgeleiding. Aan het einde van deze sessie willen we u nog een klein presentje als waardering voor uw medewerking overhandigen.

Wat houden de video-opnamen precies in?
In de praktijk zal het er op neer komen dat er bij de tweede sessie een aantal camera's op statieven worden geplaatst in de kamer om uw gezichtsuitdrukkingen te filmen. Deze camera's zullen alleen aan staan op het moment dat het interview begint.

Wat houden de audio-opnamen precies in?
Er wordt centraal tussen u en de onderzoeker een microfoon geplaatst. Daarnaast krijgen u en de onderzoeker ieder een eigen draagbare microfoon omgehangen zodat we uw stem goed kunnen opnemen.

Wat houden de metingen in?
Voor het meten van uw hartslag worden er een soort pleisters op uw huid geplakt. Deze pleisters zijn anti-allergisch en ontworpen om gemakkelijk verwijderd te worden na de studie. Voordat deze pleisters geplaatst kunnen worden moet de huid op deze plaatsen even schoongemaakt worden met een alcoholdoekje. Er zijn 2 sensoren ter hoogte van het sleutelbeen en 2 sensoren op navelhoogte. De laatste sensor zit links op uw zijkant. De afbeelding laat zien waar we een pleister plakken.



Voor het meten van de huidgeleiding worden er aan twee vingers meetapparaatjes geplaatst. Hieronder ziet u een voorbeeld hoe deze bevestigd zullen worden.

Een derde apparaatje meet hoeveel u beweegt. Deze wordt om uw pols bevestigd.
Hieronder ziet u een afbeelding hiervan.

Alle meetapparaten zullen door de onderzoekers worden bevestigd die er op zijn getraind dit snel en met zo weinig mogelijk last te verwijderen. Ze zijn ontworpen om zo min mogelijk ongemak te veroorzaken. Mocht u hier vragen over hebben, neem dan gerust contact op.

Wat zijn voordelen van deelname?
Uw deelname aan de studie wordt zeer op prijs gesteld. Het draagt bij aan belangrijk wetenschappelijk onderzoek naar hoe levensalbums emoties kunnen opwekken. Deze inzichten kunnen in de toekomst worden toegepast om de stemming van mensen te meten en met levensalbums te verbeteren.

Is mijn deelname vrijwillig?
We geven u graag volledige vrijheid om deel te nemen. Als u niet mee wilt doen, dan hoeft u daarvoor geen reden te geven. Ook als u op dit moment besluit om wel mee te doen, kunt u zich later altijd nog bedenken. Tijdens de studie mag u vragen weigeren als u zich niet comfortabel daarbij voelt of een pauze inlassen als u dat zou willen.

Wat gebeurt er met mijn gegevens?
We gaan zeer zorgvuldig met uw gegevens om omdat we snappen dat het gevoelige gegevens zijn (video en audio gegevens). Uw gegevens zullen worden beveiligd door de Universiteit Twente en zullen niet zonder uw toestemming worden gedeeld. Ook zullen uw gegevens absoluut niet voor commerciële promotiedoeleinden worden gebruikt.

Op de schriftelijke toestemming kunt u aangeven voor welke doeleinden uw gegevens mogen worden gedeeld. Deze doeleinden zijn als volgt:

We geven u de optie om aan te geven of u uw gegevens wilt delen met andere wetenschappelijke onderzoekers die uw gegevens kunnen gebruiken in hun studies. Dit zal gebeuren in de vorm van een database waarin alle gegevens van deze studie zijn opgeslagen. Deze database is beveiligd door de Universiteit Twente. Alle wetenschappelijke onderzoekers die de database willen gebruiken moeten een licentie aanvragen bij de hoofdonderzoekers waarin zij verklaren dat de gegevens niet publiek gedeeld mogen worden of voor commerciële promotiedoeleinden worden gebruikt. Ook moeten ze aangeven de gegevens niet verder te delen en alleen voor hun specifieke studie te gebruiken Na het ondertekenen van de licentie en goedkeuring van de hoofdonderzoekers, krijgen de wetenschappelijke onderzoekers pas toegang.

Ook geven we de optie om aan te geven of uw gegevens gepresenteerd mogen worden bij colleges voor studenten of bij wetenschappelijke conferenties of presentaties. Dit kan alleen als u hier toestemming voor geeft.

Alle vragenlijsten worden genummerd en apart van uw persoonlijke gegevens bewaard. Zolang het project loopt kunt u al uw gegevens bij de onderzoeker opvragen.

Ondertekenen toestemmingsverklaring
Als u besluit deel te nemen, vragen we u daarvoor schriftelijke toestemming te geven. Dit zal gebeuren tijdens de eerste kennismakingssessie. Hiermee bevestigt u uw voornemen mee te doen. Ook bevestigt u daarmee dat we u voldoende geïnformeerd hebben over de studie.

Ten slotte
Deze studie wordt uitgevoerd na een positief oordeel van de Commissie Ethiek van de Universiteit Twente te Enschede. De voor deze studie internationaal vastgestelde richtlijnen voor de studie en databescherming zullen nauwkeurig in acht worden genomen.

Verdere informatie
[…]

Na het ontvangen van deze informatiebrief zullen wij u binnen een aantal dagen bellen om te vragen of u interesse heeft om mee te doen aan deze studie. Tijdens dit telefonisch gesprek kunt u al uw vragen stellen aan de onderzoekers. Natuurlijk mag u ook eerder of juist later nog een keer bellen voor meer informatie of om aan te geven dat u wilt meedoen.

Appendix III: Instructions for annotating valence

Valence of Emotional Memories scale (VEM)

A memory is often part of a life event. A life event is a normative event that is likely to occur in the life of the average person – for example, graduation, first employment, marriage, or childbirth (Berntsen & Rubin, 2004). Consequently, a memory can be characterized as (part of) a certain life event.

Scoring valence

In the Excel file, each memory fragment has a column with the transcript of that particular fragment (Transcripts), the relevant life event (Life_Events), possible specifics of the life event if relevant (Life_Events_Specifics), a valence score (Valence_Score), a subjectivity score if relevant (Subjectivity_Score), and a reflective emotion score if relevant (Reflective_Score), as shown in Table 1. The life events, specifics, and valence scores are based on the Valence of Emotional Memories scale (VEM) depicted in Table 2.

Table 1. *Header of the Excel file for annotating valence.*

Target_Word	Emotional_ Memory	Timestamp	Events_ VEM	Specifics_ VEM	Valence_ Score	Subjectivity_ Score	Sum_ Valence	Reflective_ Score	Transcript	Notes

Instructions for annotating valence

1. Based on the VEM, establish the most fitting life event for the fragment. If the participant describes a memory that is not represented in the VEM, choose the life event **most fitting**.
2. If relevant, establish the most fitting specifics of the event (if represented in the VEM).
3. Fill in the valence score of the life event.
4. If the transcript of the life event **explicitly** states that the fragment is of a different valence than the other life events described by the participant, than a subjectivity score of +1/-1 can be given to lower or higher the original valence score of the life event. This adjustment score **can only be given** when the participant explicitly states or indicates that this particular life event is different than the other events in terms of emotion or valence.
5. The total valence score of a life event can be calculated by the sum of the valence score and the subjectivity score.
6. For certain fragments, it is possible that the participant describes emotions and/or reflections of their emotions, instead of describing a memory or life event. Only if the participant is reflective about their own **emotions**, a reflective score can be assigned. If so, a reflective score can be given by scoring a '1'. In principle, no valence score is assigned when a reflective score is assigned, **unless** the reflection appears in the same fragment as the memory. In this case, the valence score is the dominant score.
7. If a novel event is described that does not show any agreement with a life event in the VEM, a new life event can be developed by selecting multiple key words of the lexicon of Moors and colleagues (2013) that together define the novel event. The valence of this new life event is established by calculating the mean valence of the selected key words. A new life event may **only** be generated if the described event shows no connections to any other life event in the VEM.

Table 2. *Valence of Emotional Memories scale (VEM). Left: specifics are arranged on the basis of a common factor. Right: life events are sorted from low to high valence score.*

Life event	Specifics	Valence
Death	Child	1.14
	Grandchild	1.14
	Partner	1.17
	Sibling	1.33
	Parent	1.40
	Grandparent	1.60
	Friend	1.76
Neglected by children		1.50
War		1.56
First rejection		1.74
Serious disease		1.92
Divorce	Own	2.00
	Parents	2.08
Career failure		2.00
Infidelity		2.12
Financial troubles		2.21
In an accident		2.30
Operation		2.40
Family quarrels		2.45
Caring for parents		2.75
Psychological problems		3.00
Empty nest		3.83
Prepare for death		4.13
Move		4.13
Other child's milestones		4.20
Puberty		4.68
Military service		4.85
Hobby		4.86
Retirement		4.97
Youth		5.00
Leave home		5.22
Begin school		5.53
High school		5.55
First sexual experience		5.64
College		5.88
Birth	Sibling	5.40
	Child	6.34
	Grandchild	6.66
Achievement	First job	5.48
	Career success	5.94
	Buying a house	6.06
	Driver's license	6.24
	Big achievement	6.28
	The 'right' job	6.37
	Graduation	6.54
Travelling		6.14
Having friends		6.30

Life event	Valence
Child's death	1.14
Grandchild's death	1.14
Partner's death	1.17
Sibling's death	1.33
Parent's death	1.40
Neglected by children	1.50
War	1.56
Grandparent's death	1.60
First rejection	1.74
Friend's death	1.76
Serious disease	1.92
Own divorce	2.00
Career failure	2.00
Parents' divorce	2.08
Infidelity	2.12
Financial troubles	2.21
In an accident	2.30
Operation	2.40
Family quarrels	2.45
Caring for parents	2.75
Psychological problems	3.00
Empty nest	3.83
Prepare for death	4.13
Move	4.13
Other child's milestones	4.20
Puberty	4.68
Military service	4.85
Hobby	4.86
Retirement	4.97
Youth	5.00
Leave home	5.22
Sibling's birth	5.40
First job	5.48
Begin school	5.53
High school	5.55
First sexual experience	5.64
College	5.88
Career success	5.94
Buying a house	6.06
Travelling	6.14
Driver's license	6.24
Big achievement	6.28
Having friends	6.30
Own marriage	6.33
Child's birth	6.34
Falling in love	6.34
The 'right' job	6.37

Table 2. *Continued.*

Life event	**Specifics**	**Valence**
Marriage	Own	6.33
	Child	6.47
Falling in love		6.34
Child's college graduation		6.66
Family holidays		6.67
Celebrations		6.75

Life event	**Valence**
Child's marriage	6.47
Graduation	6.54
Child's college graduation	6.66
Grandchild's birth	6.66
Family holidays	6.67
Celebrations	6.75

Appendix IV: Script for sentiment analysis

```
#!/usr/bin/python3
# geschreven door Judith van Stegeren and Lorenzo Gatti
# j.e.vanstegeren@utwente.nl ; l.gatti@utwente.nl
# 12 maart 2019


#@author: Deniece Nazareth
"""

#from pattern.nl import singularize, sentiment #original pattern.nl
#from pattern.nl_moors import singularize, sentiment #pattern + moors
from pattern.nl_moors_subset import singularize, sentiment #pattern + subset of moors where
word intensity >= 0.3
from nltk import sent_tokenize
import csv
import pandas as pd
import os
import matplotlib.pyplot as plt
import statistics
from openpyxl import load_workbook
import numpy as np
from scipy import stats

#Open file
filename =
"/Users/Deniece/Desktop/S1_12_Sentiment/P23_Timestamps_Events_Rater2.xlsx"


#skip the two first rows that are empty. 3rd row is header and we want to keep it!
rows_to_skip = 1


xl = pd.read_excel(io=filename,sheetname=1,header=1,skiprows=range(0,rows_to_skip-1),
data_only = True)
#Drop the Reflective_Emotions as they do not have a valence_Score and are therefore not
relevant
xl= xl.dropna(subset = ["valence_Score"])

xl["Transcripts"].fillna(",inplace=True)


##### CALCULATING THE SENTIMENT #####

#Create empty lists for the sentiment per sentence, per event for all sentences wrt mean and
std,per words in an event to append later
sentiment_per_sentence = []
sentiment_per_event = []
sentiment_event_avg_sentences = []
```

```
#sentiment_event_zscore_sentences = []
sentiment_event_std_sentences = []
sentiment_per_word_event = []

#first for loop to calucalte polarity scores for each submemory/timestamp. 2nd for loop to
calculate
#polarity/sentiment scores and assessment scores for each sentence in the submemories

for m in xl["Transcripts"]:
  if not m:
        continue
  senti_memory_score = sentiment (m)
  print (senti_memory_score)
  print (sentiment(m).assessments)
  #sentiment_per_event.append(senti_memory_score[0])
  tmp_sent_sentiment_per_sentence = []
  sentiment_per_word =[]
  for s in sent_tokenize(m):
    print("Sentence:",s)
    senti_score = sentiment(s)
    senti_score_assessments = sentiment(s).assessments
    print("Sentiment score (polarity, subjectivity):",senti_score)
    print("Judgement based on words:",senti_score_assessments)
    sentiment_per_sentence.append(senti_score[0])
    tmp_sent_sentiment_per_sentence.append(senti_score[0])
    sentiment_per_word.append(senti_score_assessments)
    print()
  sent_dev = 0
  if len(tmp_sent_sentiment_per_sentence)>1:
    sent_dev = statistics.stdev(tmp_sent_sentiment_per_sentence)
  sentiment_event_avg_sentences.append(statistics.mean(tmp_sent_sentiment_per_sentence))
  #sentiment_event_zscore_sentences.append(stats.zscore(tmp_sent_sentiment_per_sentence))
  sentiment_event_std_sentences.append(sent_dev)
  words_in_cell = ""
  for sentence in sentiment_per_word:
    words_per_sentence = ", ".join([" ".join([token for token in words[0]])+"
"+'{:.3f}'.format(words[1]) for words in sentence])
    words_in_cell = words_per_sentence + " # " + words_in_cell
  sentiment_per_word_event.append(words_in_cell)


##### SAVING MEAN AND STD TO NEW SHEET IN EXISTING EXCIL FILE #####


#create a new dataframe for storing the average of the sentiment of sentences per event (and
the stdev too)
new_worksheet_content = pd.DataFrame()
new_worksheet_content["sentiment_sentences_avg"] = sentiment_event_avg_sentences
new_worksheet_content["sentiment_sentences_stdev"] = sentiment_event_std_sentences
```

```
#new_worksheet_content["sentiment_event_zscore_sentences"] =
sentiment_event_zscore_sentences
new_worksheet_content["sentiment_per_word_event"] = sentiment_per_word_event


#reload the excel file using the openpyxl engine
book = load_workbook(filename)
writer = pd.ExcelWriter(filename, engine='openpyxl')
writer.book = book
#copy-paste all the existing worksheets so the file is not modified
writer.sheets = dict((ws.title, ws) for ws in book.worksheets)

#add a new worksheet with name "sentiment" containing our data
new_worksheet_content.to_excel(writer, "sentiment")

#save the file
writer.save()
```

Appendix V: Praat script for extracting acoustic features

```
## start procedure De Jong for syllable detection ##

procedure countSyll

# shorten variables
#iglevel = 'ignorance_Level/Intensity_Median'
#mindip = 'minimum_dip_between_peaks'

iglevel = 0
mindip = 2

obj$ = selected$("Sound")
soundid = selected("Sound")
originaldur = Get total duration

Subtract mean

# Use intensity to get threshold
To Intensity... 50 0 yes

intid = selected("Intensity")

start = Get time from frame number... 1

nframes = Get number of frames
end = Get time from frame number... 'nframes'

# estimate noise floor
minint = Get minimum... 0 0 Parabolic

# estimate noise max
maxint = Get maximum... 0 0 Parabolic

#get median of Intensity: limits influence of high peaks
medint = Get quantile... 0 0 0.5

# estimate Intensity threshold
threshold = medint + iglevel
if threshold < minint
  threshold = minint
endif

Down to Matrix

matid = selected("Matrix")
# Convert intensity to sound
To Sound (slice)... 1
```

```
sndintid = selected("Sound")

intdur = Get finishing time
intmax = Get maximum... 0 0 Parabolic

# estimate peak positions (all peaks)
To PointProcess (extrema)... Left yes no Sinc70

ppid = selected("PointProcess")

numpeaks = Get number of points

# fill array with time points
for i from 1 to numpeaks
  t'i' = Get time from index... 'i'
endfor

# fill array with intensity values

select 'sndintid'

peakcount = 0
for i from 1 to numpeaks
  value = Get value at time... t'i' Cubic
  if value > threshold
    peakcount += 1
    int'peakcount' = value
    timepeaks'peakcount' = t'i'
  endif
endfor

# fill array with valid peaks: only intensity values if preceding
# dip in intensity is greater than mindip

select 'intid'

validpeakcount = 0
precedingtime = timepeaks1
precedingint = int1
for p to peakcount-1
  following = p + 1
  followingtime = timepeaks'following'
  dip = Get minimum... 'precedingtime' 'followingtime' None
  diffint = abs(precedingint - dip)
  if diffint > mindip
    validpeakcount += 1
    validtime'validpeakcount' = timepeaks'p'
  endif
  precedingtime = timepeaks'following'
  precedingint = Get value at time... timepeaks'following' Cubic
```

```
endfor

# Look for only voiced parts

select 'soundid'

To Pitch (ac)... 0.02 30 4 no 0.03 0.25 0.01 0.35 0.25 450
pitchid = selected("Pitch")

voicedcount = 0
for i from 1 to validpeakcount
  querytime = validtime'i'

  value = Get value at time... 'querytime' Hertz Linear

  if value <> undefined
    voicedcount = voicedcount + 1
    voicedpeak'voicedcount' = validtime'i'
  endif
endfor

# calculate time correction due to shift in time for Sound object versus
# intensity object
timecorrection = originaldur/intdur

# Insert voiced peaks in second Tier

select 'soundid'

To TextGrid... "syllables" syllables
textgridid = selected("TextGrid")
for i from 1 to voicedcount
  position = voicedpeak'i' * timecorrection
  Insert point... 1 position 'i'
endfor

#select TextGrid 'textgridid'
n_syllables = Get number of points... 1
#printline 'n_syllables'
.result = n_syllables

# write textgrid to textfile
#Write to text file... 'directory$'/'obj$'.syllables.TextGrid
#Write to text file... /Users/khiet/work/deniece/test/'obj$'.syllables.TextGrid


# clean up before next sound file is opened

select 'intid'
plus 'matid'
```

```
plus 'ppid'
plus 'sndintid'
plus 'soundid'
plus 'textgridid'
plus 'pitchid'
Remove

endproc

## end procedure syllable detection by De Jong ##


## start script to extract features ##

 #praat  bin/pitch-stuff-perfile.praat  /Users/khiet/work/deniece/Sessie1_Sarah/P1_S1_16k.wav
/Users/khiet/work/deniece/Sessie1_Sarah/P1_finished_new.TextGrid
/Users/khiet/work/deniece/Sessie1_Sarah_sil_praat/P1_S1_16k.silpraat.TextGrid male
 #                    praat                    bin/extract-stuff-perfile.praat
/Users/khiet/work/deniece/Sessie1_Sarah/P1_S1_16k.wav
/Users/khiet/work/deniece/Sessie1_Sarah_lifeevents2/P1_Sil_Timetamps_merged.TextGrid
/Users/khiet/work/deniece/Sessie1_Sarah_sil_praat2/P1.silpraat2.TextGrid

form Get arguments
  sentence Wavfile
  sentence Tgfile
  sentence Tgfilesil
  sentence Gender
endform

myobject$ = "dummy"
myobject_sil$ = "sil"
Read from file... 'wavfile$'
name_wavfile$ = selected$("Sound")
where = index(name_wavfile$,"_")
name_id$ = left$(name_wavfile$,where-1)
number_id$ = mid$(name_wavfile$,2,where-2)
Rename... 'myobject$'

Read from file... 'tgfilesil$'
Rename... 'myobject_sil$'

Read from file... 'tgfile$'
Rename... 'myobject$'

select TextGrid 'myobject$'
n_intervals = Get number of intervals... 1

step = 0.01
t = step
begin_low_band = 0
```

```
end_low_band = 2000
begin_high_band = 2000
end_high_band = 5000

if (gender$ = "male")
  pitch_range = 300
  number_gender = 0
else
  pitch_range = 600
  number_gender = 1
endif



## header
#print
speaker_id1;speaker_id2;gender;label;interval_number;starttime;endtime;dur_part;mean_pitc
h;sd_pitch;min_pitch;max_pitch;range_pitch;slope_pitch;
#print mean_intens;sd_intens;min_intens;max_intens;range_intens;
#print mean_intens_all;sd_intens_all;min_intens_all;max_intens_all;range_intens_all;
#print hammarberg;slope_ltas;cog;
#print n_syllables;ar;sr;
#print mean_dur_silent;mean_dur_talkspurt;ratio_dur_sil_talkspurt;talkspurt_rate;sil_rate
#print 'newline$'

for j from 1 to n_intervals
  select TextGrid 'myobject$'
  label$ = Get label of interval... 1 'j'
  where = startsWith(label$,"Timestamp")
  if (where==1)

    starttime = Get start time of interval... 1 'j'
    endtime = Get end time of interval... 1 'j'
    select Sound 'myobject$'
    Extract part... 'starttime' 'endtime' rectangular 1 no
    dur_part = Get total duration
    select TextGrid 'myobject_sil$'
    Extract part... 'starttime' 'endtime' no
    select Sound 'myobject$'_part
    plus TextGrid 'myobject_sil$'_part
    Extract intervals where... 1 no "is equal to" sounding
    Concatenate
    dur_chain = Get total duration

    ## do stuff here

    ## pitch stuff
    # set 500 for female voices
    # set 300 for male voices
    To Pitch... 0.0 75 'pitch_range'
```

```
    mean_p = Get mean... 0 0 Hertz
    sd_p = Get standard deviation... 0 0 Hertz
    # min = 5th percentile
    # max = 95th percentile
    min_p = Get quantile... 0 0 0.05 Hertz
    max_p = Get quantile... 0 0 0.95 Hertz
    range_p = max_p - min_p
    slope_p = Get mean absolute slope... Hertz

    print
'"'name_id$'"';'number_id$';'number_gender';'"'label$'"';'j';'starttime:2';'endtime:2';'dur_part:2';'
mean_p:4';'sd_p:4';'min_p:4';'max_p:4';'range_p:4';'slope_p:4';

    ## intensity stuff over only speech part
    select Sound chain
    To Intensity... 100 0 yes
    mean_intens = Get mean... 0 0 energy
    sd_intens = Get standard deviation... 0 0
    # min = 5th percentile
    # max = 95th percentile
    min_intens = Get quantile... 0 0 0.05
    max_intens = Get quantile... 0 0 0.95
    range_intens = max_intens - min_intens

    print 'mean_intens:4';'sd_intens:4';'min_intens:4';'max_intens:4';'range_intens:4';

    ## intensity stuff over whole part
    select Sound 'myobject$'_part
    To Intensity... 100 0 yes
    mean_intens_all = Get mean... 0 0 energy
    sd_intens_all = Get standard deviation... 0 0
    # min = 5th percentile
    # max = 95th percentile
    min_intens_all = Get quantile... 0 0 0.05
    max_intens_all = Get quantile... 0 0 0.95
    range_intens_all = max_intens_all - min_intens_all

    print
'mean_intens_all:4';'sd_intens_all:4';'min_intens_all:4';'max_intens_all:4';'range_intens_all:4';

    ## voice quality stuff
    select Sound chain
    To Ltas... 100
    max_low_band = Get maximum... 'begin_low_band' 'end_low_band' None
    max_high_band = Get maximum... 'begin_high_band' 'end_high_band' None
    hammarberg = max_low_band - max_high_band
    slope_ltas = Get slope... 'begin_low_band' 'end_low_band' 'begin_high_band'
'end_high_band' energy

    select Sound chain
```

```
    To Spectrum... yes
    cog = Get centre of gravity... 2

    print 'hammarberg:4';'slope_ltas:4';'cog:4';

    #Save as WAV file... /Users/khiet/work/deniece/test/chain.syllables.wav
    ## count syllables
    select Sound chain
    @countSyll
    ar = countSyll.result/dur_chain
    sr = countSyll.result/dur_part
    print 'countSyll.result';'ar:4';'sr:4';

    ## pause stuff
    select TextGrid 'myobject_sil$'_part
    tot_dur_silent = Get total duration of intervals where... 1 "is equal to" silent
    tot_dur_talkspurt = Get total duration of intervals where... 1 "is equal to" sounding
    n_silent = Count intervals where... 1 "is equal to" silent
    n_talkspurt = Count intervals where... 1 "is equal to" sounding
    if (n_silent > 0)
      mean_dur_silent = tot_dur_silent/n_silent
    else
      mean_dur_silent = 0
    endif
    mean_dur_talkspurt = tot_dur_talkspurt/n_talkspurt
    ratio_dur_sil_talkspurt = tot_dur_silent/tot_dur_talkspurt
    talkspurt_rate = n_talkspurt/dur_part
    sil_rate = n_silent/dur_part

    print
'mean_dur_silent:4';'mean_dur_talkspurt:4';'ratio_dur_sil_talkspurt:4';'talkspurt_rate:4';'sil_rat
e:4'

    print 'newline$'

    select Sound 'myobject$'_part
    plus Pitch chain
   #plus Sound chain
    plus Ltas chain
    plus Intensity chain
    plus Intensity 'myobject$'_part
    plus Spectrum chain
    plus TextGrid 'myobject_sil$'_part

    Remove

  endif


endfor
```

Appendix VI: Acoustic speech features extracted with Praat

Acoustic feature	**Description**
Mean_pitch	Averaged pitch over non-silenced parts
Sd_pitch	Standard deviated pitch over non-silenced parts
Min_pitch	Minimum pitch (5th percentile of pitch values)
Max_pitch	Maximum pitch (95th percentile of pitch values)
Range_pitch	Pitch range (maximum pitch – minimum pitch)
Slope_pitch	Mean absolute pitch slope in Hz
Mean_intens	Averaged intensity over non-silenced parts
Sd_intens	Standard deviated intensity over non-silenced parts
Min_intens	Minimum intensity (5th percentile of intensity values)
Max_intens	Maximum intensity (95th percentile of intensity values)
Range_intens	Intensity range (maximum intensity – minimum intensity)
Mean_intens_all	Averaged intensity over silenced and non-silenced parts
Sd_intens_all	Standard deviated intensity over silenced and non-silenced parts
Min_intens_all	Minimum intensity (5th percentile of intensity values)
Max_intens_all	Maximum intensity (95th percentile of intensity values)
Range_intens_all	Intensity range (maximum intensity – minimum intensity)
Hammarberg	Hammarberg Index
Slope_ltas	Slope over the low (0-2000 Hz) and high band (2000-5000 Hz)
Cog	Spectral center of gravity
N_syllables	Number of syllables
Ar	Articulation rate
Sr	Speech rate
Mean_dur_silent	Mean duration of silent parts
Mean_dur_talkspurt	Mean duration of sounding parts
Ratio_dur_sil_talkspurt	Ratio of duration of silent/sounding parts
Talkspurt_rate	Sounding rate
Sil_rate	Silent rate