Voices of Anxiety: Do Voice Dynamics Reveal Anxiety Levels?

Gerrieke Druijff-van de Woestijne

Master Science Thesis

Supervised by:
1. Anna Lichtwarck-Aschoff
2. Fred Hasselman

June 30, 2018

Abstract

Understanding how anxious children change during treatment requires anxiety measurements. Since self-reported questionnaires for children are suboptimal, unobtrusive alternatives are examined. Both rigidity in physiological signals (e.g. heart rate) and changes in voice features have been related to anxiety. We investigated whether voice dynamics reflected children's anxiety levels. Audio recordings of therapy sessions with children with anxiety (n = 41) and of conversations with control children (n = 41) were used. From every therapy session and conversation, audio segments about fear-related and neutral topics were selected. Children's voices were analyzed with recurrence quantification analyses, providing information about structure and patterns in the signal. In general, control children had voices with lower recurrence measures than anxious children. Children who responded to treatment showed higher overall recurrence measures, but voice dynamics did not change over the course of treatment. Higher recurrence measures here indicate both high regularity and low predictability. These findings are difficult to interpret in light of earlier studies and express the need for research in more controlled settings to understand the relationship between voice dynamics and anxiety.

Voices of Anxiety: Do Voice Dynamics Reveal Anxiety Levels?

Childhood anxiety is a prevalent psychiatric disorder with consequences lasting in adulthood if untreated (Copeland, et al., 2013; Kessler, Petukhova, Sampson, Zaslavsky, & Wittchen, 2012). Cognitive-behavioral therapy (CBT), the best evidence-based treatment, is estimated to be successful in approximately sixty percent of cases (James, James, Cowdrey, Soler, & Choke, 2015). Monitoring children in treatment could allow for early observation of the treatment's effectiveness, making it possible to stop ineffective treatments early. For example, Bettis, Forehand, Sterba, Preacher and Compas (2016) were able to predict treatment response based on monitored anxiety levels.

However, monitoring children's anxiety levels with the use of self-reports may not be an optimal approach. Especially for young children, understanding and reliably answering questions regarding inner states can be difficult (Safford, Kendall, Flannery-Schroeder, Webb, & Sommer, 2005; Schwarz, 1999). Monitoring also requires repeated and frequent measurements, which could put additional burden on clients (Shiffman, Stone, & Hufford, 2008; Wenze & Miller, 2010).

These difficulties regarding self-reports have made the need for more automatic and unobtrusive measures more apparent. For this purpose, physiological measures have been used in earlier studies. Anxiety can be regarded a state of less adaptive and more rigid behavior, which can be revealed in physiological measures like heart rate. Reductions in heart rate complexity are associated with a range of poor health outcomes (Friedman & Thayer, 1998). Adolescents with high levels of anxiety have lower heart rate complexity than adolescents with low levels of anxiety (Bornas, Balle, De la Torre-Luque, Fiol-Veny, & Llabres, 2015). Across a variety of anxiety disorders, heart rate variability (indicating flexibility or adaptation) is lower than in healthy controls (Pittig, Arch, Lam, & Craske, 2013), and extremely shy children at risk for adult anxiety disorders have more stable heart rates (Friedman & Thayer, 1998). Supposedly, it would be possible to measure heart rate over the course of treatment to monitor anxiety levels. However, this could be inconvenient for both client and therapist.

However, there is another physiological signal that has been associated with anxiety, namely speech. Therapy sessions are made up of speech, making it easy to assess this variable. People exchange information vocally not only in terms of content (i.e. the topic of the conversation) but also of form (i.e. the vocal characteristics). For example, people with quiet voices are considered timid or shy in daily life. To quantify speech characteristics, amplitude and fundamental frequency can be extracted, indicators of loudness and pitch of the voice, respectively. Fundamental frequency has been described as vocally encoded arousal or emotional state and has been associated with other physiological measures, such as heart rate, blood pressure, and cortisol (Bugental, Beaulieu, Schwartz, & Dragosits, 2009; Imel, et al., 2014).

Several studies investigated how speech of children with anxiety differs from healthy controls, mostly with a focus on social anxiety. With regard to pitch, children with social anxiety had higher pitch and more restricted pitch ranges than typically developing children (Kroytor, 2012; Scharfstein & Beidel, 2014; Scharfstein, Beidel, Sims, & Finnell, 2011). Children with generalized anxiety were not significantly different from typically developing children (Scharfstein & Beidel, 2014). With regard to loudness, children with social anxiety had lower volume and less volume variability than typically developing children (Kroytor, 2012; Scharfstein & Beidel, 2014; Scharfstein, et al., 2011). Again, children with generalized anxiety did not significantly differ from typically developing children (Scharfstein & Beidel, 2014). Over the course of treatment, volume of children with social anxiety increased, so that they were similar to typically developing children (Kroytor, 2012). However, this effect should be taken with caution because it was reported in an archived bachelor thesis that also had contradictory statements in the results section.

For adults with anxiety, similar effects were reported. Changes in pitch and loudness following a social threat differed between people with high and low levels of social anxiety (Gilboa-Schechtman, Galili, Sahar, & Amir, 2014). People with social anxiety disorder had higher pitch during social threat but not during a diagnostic interview than healthy controls (Weeks, et al., 2012). However, pitch of males with social anxiety was also higher at the start of the diagnostic interview and associated with symptom severity in another study (Weeks, Srivastav, Howell, & Menatti, 2016). Pitch of neutral sentences was positively associated with symptoms of social anxiety in a community sample (Galili, Amir, & Gilboa-Schechtman, 2013), but only for males in an earlier study (Weeks, et al., 2012). No effects on variability of pitch and loudness or effects over time were reported in these studies. For those who responded to a pharmacological treatment, pitch and maximal pitch decreased from baseline to post-treatment, which was also associated with state anxiety (Laukka, et al., 2008). Differences in voice characteristics in response to fear-related threat were also found in a study on specific phobia. Women with high fear of snakes or spiders showed increased voice volume in short responses ('up' or 'down') to pictures of feared compared to non-feared animals (Flykt, Bänziger, & Lindeberg, 2017).

Thus, pitch and loudness seem to be related to anxiety levels in general and clinical populations, for both adults and children. Whether these voice characteristics are sensitive to improvement due to treatment is less clear, since only few studies investigated these effects. Studies mentioned so far mainly quantified voice characteristics using mean fundamental frequency and mean amplitude. Although these averaged variables give basic information about voices, much information is lost. Speech production is the result of cooperation between approximately eighty different muscles, making it a complex system that is better described by other indicators (Turvey, 2007). Incorporating ranges and variability, next to mean values, provides a more nuanced picture but still neglects the temporal order within the signal. Each individual voice has its unique timbre, that is defined by more subtle characteristics (shapes of the signal) than loudness and fundamental frequency. In fact, recognition of voices, music instruments, and words requires these more advanced features. A possibility to analyze voices retaining the richness and temporal order in the signal is trough recurrence quantification analysis (RQA), that has been used to identify words (Rufiner, Torres, Gamero, & Milone, 2004) and effects of stress on speech (Jackson, Tiede, Beal, & Whalen, 2016). RQA can be used to identify temporal dynamics, based on patterns and structures in the signal (Zbilut & Webber, 1992).

Using this technique for speech, voice disorders could be classified with high accuracy (Little, McSharry, Roberts, Costello, & Moroz, 2007). Fusaroli and colleagues managed to classify mental disorders using RQA of speech signals. Temporal dynamics of freely spoken descriptions of videos with moving shapes distinguished between children with autism spectrum disorder and controls (Fusaroli, Grossman, Cantio, Bilenberg, & Weed, 2015). Using the same methods, speech of patients with schizophrenia, Asperger's, depression, and healthy controls could be distinguished (Fusaroli, Tylén, Simonsen, & Weed, 2013).

Given its promising results regarding a variety of mental disorders, RQA may also be suited to find voice differences that are associated with anxiety. Other research on voice characteristics found that pitch and volume can be indicators of anxiety and that these features may also be indicative of improvement due to treatments(e.g. Kroytor, 2012; Scharfstein, et al., 2011). However, these studies did not take into account that voices are complex signals, which can be accounted for by RQA. Using RQA measures for voice features could be a new way to monitor anxiety levels.

Based on these results, the current study tries to investigate whether children’s speech signal conveys information about anxiety. We pose the following questions: 1) Are anxiety levels related to voice characteristics and 2) how do these voice characteristics change over treatment? To answer these questions, we compared speech of children who received anxiety treatment to speech of control children. Within the treatment group, we also compared improvers to non-improvers. Last, speech covering fear-related topics was compared to speech covering neutral topics. RQA’s were used to quantify voice characteristics.

Method

Participants

Data from an earlier study investigating the effectiveness of CBT was used (Janssen, et al., 2012; Van Doorn, Jansen, Bodden, Lichtwarck-Aschoff, & Granic, 2017). In this effectiveness study, a manualized CBT treatment for childhood anxiety problems was compared to treatment-as-usual, in two Dutch mental health care institutions. Children whose anxiety levels (either indicated by themselves or their mothers) exceeded normal range on the Screen for Child Anxiety Related Emotional Disorders (SCARED; Birmaher, et al., 1999) Total scale or one of four subscales (Separation anxiety, Social anxiety, Generalized anxiety, Panic disorder) were eligible for inclusion. Children with posttraumatic stress disorder, autism spectrum disorder, specific phobia, obsessive-compulsive disorder, or an IQ below 80, and children that needed immediate intervention to prevent harm to themselves or their family were excluded. The 88 children that participated in the effectiveness study formed the treatment group in the current study. These were clinically anxious children aged between 7 and 12 who had been referred to mental health care centers and were randomly assigned to either CBT (n = 43) or treatment-as-usual (n = 45). Due to time restrictions, only the first 44 children of the treatment group could be included in this study. Children for whom no audiotaped therapy sessions were available were excluded, leaving 41 children (20 CBT, 21 treatment-as-usual).

In addition, a control group (n = 46) was recruited that consisted of children aged between seven and eleven years who did not concurrently receive treatment for anxiety problems at the time. These children were recruited via a primary school using a sampling strategy to ensure that the distribution of ages and gender approximated that of the treatment group. Children with high anxiety levels, measured by the SCARED (Birmaher, et al., 1999) were excluded from the control group, leaving 41 children.

Ethical approval for both the effectiveness study (ECG16122010) and the current study including data collection of the control group (ECSW-2017-06) was granted by the ethical committee of the Faculty of Social Sciences at Radboud University Nijmegen.

Procedure

Treatment group. When parents gave written consent, children were randomly assigned to CBT or treatment-as-usual. Therapy sessions were audiotaped by therapists and weekly assessments of anxiety level were conducted by means of telephone calls by research assistants during three months of treatment. Before treatment, after treatment, at six months follow up and at twelve months follow up, assessments of anxiety, problem behavior, therapeutic alliance and parenting were conducted. Parents received a financial contribution and children received little gifts to compensate for participation.

Control group. When parents gave written consent, children were scheduled to participate during school days. Children first filled out the SCARED-NL and then had a ten minutes audiotaped conversation with a researcher. During this conversation, children talked about things that scared them or made them feel uneasy and about hobbies or games they liked. Children received little gifts to compensate for participation.

Treatment

Children that were assigned to the CBT treatment received manualized treatment according to the ‘Thinking + Doing = Daring’ protocol by Bögels (2008). Children in the treatment-as-usual group received various types of therapy including CBT techniques. Children in the control group did not receive any treatment.

CBT. Children in the CBT group received 12 weekly individual sessions of CBT and their parents three sessions according to a treatment manual by Bögels (2008). Eight therapists with a mean age of 52.63 (*SD* = 9.38) and, on average, 19.75 (*SD* = 7.59) years of experience were involved. Therapy concentrated on psycho education, fear registration, cognitive restructuring, learning coping skills, exposure in vivo, rewards and reinforcements, behavioral experiments, and relapse prevention.

Treatment-as-usual. Children in the treatment-as-usual group received the treatment that their therapists viewed as the most appropriate. Seven therapists with a mean age of 3314 (*SD* = 7.52) and, on average, 9.00 (*SD* = 5.89) years of experience were involved. During the study, three therapists changed jobs and were replaced by three new therapists. In 96% of cases, treatment consisted of CBT components augmented with eclectic therapy components.

Materials

Screen for Child Anxiety Related Emotional Disorders (SCARED-NL). The Dutch version of the SCARED was used to assess anxiety levels. This questionnaire consists of 69 questions with three answer possibilities: never/almost never, sometimes, and often (Muris, Bodden, Hale, Birmaher, & Mahyer, 2007). Sumscores are calculated for the subscales and the total scale. SCARED-NL has good validity compared to diagnoses or structured diagnostic interviews (Bodden, Bögels, & Muris, 2009). In the current study, internal consistency of child-reported anxiety levels was excellent for the Total scale (Cronbach’s α = .93); good for Social anxiety (Cronbach’s α = .86) and Panic disorder (Cronbach’s α = .83); and acceptable for Separation anxiety (Cronbach’s α = .75) and Generalized anxiety

(Cronbach's α = .79). Internal consistency of parent-reported anxiety (only available in the therapy group), was good for the Total scale (Cronbach's α = .87), Generalized anxiety (Cronbach's α = .82) and Social anxiety (Cronbach's α = .81); acceptable for Panic disorder (Cronbach's α = .76); and poor for separation anxiety (Cronbach's α = .54).

Recording. Conversations between child and therapist or child and researcher were audiotaped by therapist and researcher respectively. Recordings were made with a Sony Digital Voice Recorder (ICD-AX412F) that was placed on a table or carried along.

Data preprocessing

Audio selection. From each therapy session and conversation two audio segments with a maximum length of five seconds were selected using the Adobe Audition audio editing program. Segments were selected based on conversation topics: one covered a discussion about something the child feared whereas the other covered a neutral or slightly positive topic. The search for fear-related and neutral fragments followed a standardized procedure (e.g. neutral fragments were sought firstly in the last five minutes of the therapy session).

Audio preprocessing. Preprocessing[1] was done in R (version 3.5.0; R Core Team, 2018). Audio segments were filtered with a high-pass filter at 100 Hz using the signal package (version 0.7-6; signal developers, 2013) and then downsampled from 44.1 kHz to 2000 Hz using the tuneR package (version 1.3.2; Ligges, Krey, Mersmann, & Schnackenberg, 2016) reducing the size of objects to make it possible to analyze the fragments. A Huang-Hilbert transformation implemented in the EMD package (version 1.5.7; Kim & Oh, 2014) was applied to the audio signals, returning the instantaneous amplitude and frequency of the analytic signal. This transformation emphasizes the local characteristics of the audio signal (Huang, et al., 1996). For each segment, amplitudes and frequencies were scaled to range

[1] Script for preprocessing and RQA can be found in Appendix A

between 0 and 1. Figure 1 illustrates an audio segment and the results of its Huang-Hilbert transformation.

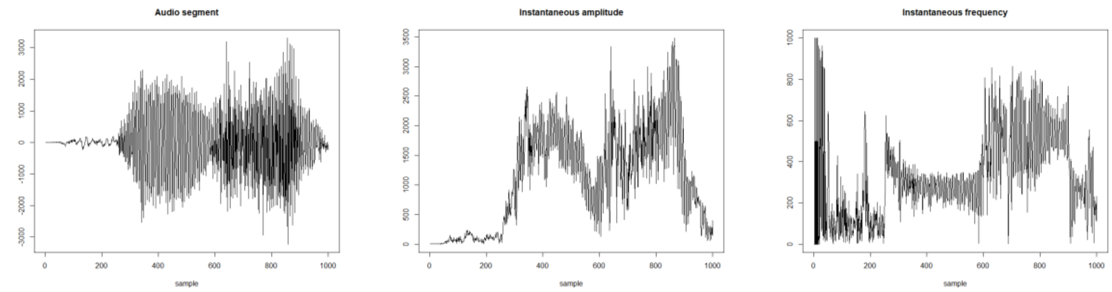

Figure 1. Example of a filtered and down-sampled voice signal (a) and the instantaneous amplitude (b) and instantaneous frequency (c) that resulted from the Huang-Hilbert transformation of this signal. This figure also shows that averaging these kind of signals causes a loss of information.

Recurrence Quantification Analysis

RQA looks for recurrent points in time series to capture temporal dynamics. It can be used for complex signals, because no assumptions regarding linearity, independence or normality of the data are made. Initially developed for physiology (Eckman, Kamphorst, & Ruelle, 1987; Zbilut & Webber, 1992), RQA was rapidly introduced unto biomedics, sociology, and behavioral sciences (Marwan, Romano, Thiel, & Kurths, 2007). Visual inspection of structures in the data is facilitated by recurrence plots, whereas quantification is possible with RQA measures, that provide information regarding randomness and predictability of the data. Marwan's procedures for RQA (http://www.recurrence-plot.tk/; Marwan, et al., 2007) that have been implemented for R in the casnet package (Hasselman, 2018) were used.

Phase space reconstruction. When only one observable is measured in discrete time it is necessary to reconstruct a phase space that describes all possible states of a system (Marwan, et al., 2007; Takens, 1981). The time delay method was used to reconstruct the phase space, which incorporates both a time delay (the lag at which components are most independent) and a number of dimensions (the number of ordinary linear equations needed to describe the system). The time delay was determined based on mutual information criteria. The first minimum for mutual information indicated that a time delay of 4 would be appropriate (based on mean and median delays) for both frequency and amplitude. Using this time delay, the number of dimensions was estimated for each frequency and amplitude time series separately using false nearest neighbors analysis. The appropriate number of dimensions ranged from 8 till 14, with 12 and 10 dimensions most frequent for amplitude and frequency, respectively. Some frequency time series were excluded from further analyses, since their number of dimensions could not be estimated. Based on a time delay of 4 and the corresponding number of dimensions estimated, the phase space was reconstructed for each audio fragment.

Recurrence quantification. To quantify the recurrences of the system, a recurrence matrix (or its visual equivalent; the recurrence plot) is defined. For a continuous variable, a radius has to be set within which points are considered recurrent. For each frequency and amplitude time series, a radius was chosen that yielded a recurrence rate of 1% (based on Marwan, et al., 2007). See Figure 2 for an illustration of a recurrence plot. From these plots or matrices, RQA measures are calculated that show characteristics of the signal.

The following measures were calculated using the function crqa_cl from the casnet package (Hasselman, 2018): DET, LAM, L_entr, and V_entr. DET is a measure of randomness, with lower DET indicating more random behavior. DET relates to the percentage of recurrent points forming diagonal lines in the recurrence plot. LAM indicates another type of randomness and relates to the number of recurrent points forming vertical lines. L_entr is a measure of the predictability of the data, with lower L_entr indicating more predictable behavior. L_entr relates to the Shannon entropy of the distribution of diagonal line lengths in the recurrence plot. V_entr indicates another type of predictability and relates to the Shannon entropy of the distribution of vertical line lengths. Thus, DET and L_entr relate to diagonal lines, which indicate that sequences of values reoccur in the signal, whereas LAM en V_entr relate to vertical lines, which indicate that a value remains constant in the signal.

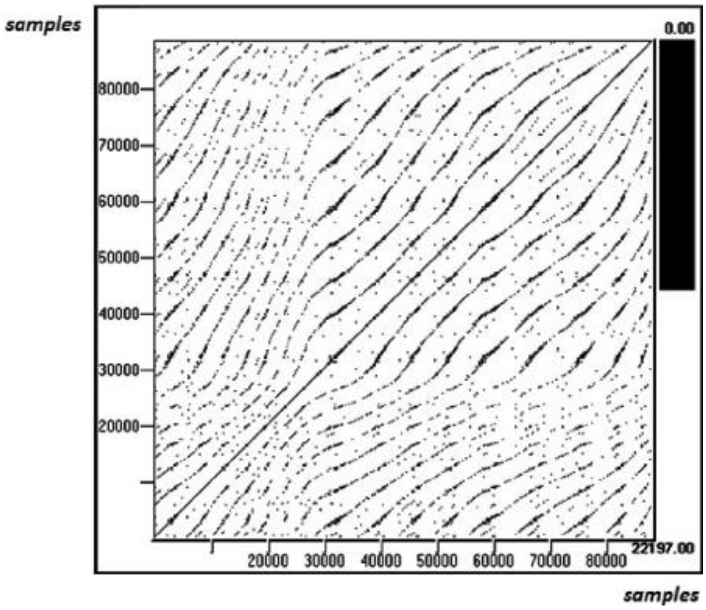

Figure 2. Example of a RQA of a child's voice, reprinted from Lopes, Costa, de Almeida Costa, Correia, and Vieira (2014).

Data-analysis

Improvers' status was defined as a function of SCARED-NL post-treatment and pre-treatment scores. Since inclusion in the treatment group of the study was based on SCARED-NL subscales, we used personalized anxiety scores as an outcome measure. The highest standardized SCARED-NL subscale at pre-treatment, either reported by mother or child, was taken as a measure of effect. Children who fell in the normal range of that subscale at post-treatment were considered improved. Children whose scores did not return within the normal range, were considered non-improved.

Mixed-effects models from the lme4 package (version 1.1.17; Bates, Maechler, Bolker, & Walker, 2015) for R were used to account for individual differences. For all analysis, numerical variables were centered at zero and contrasts for categorical variables were set to sum-to-zero coding. With regards to random effects, a maximal model as proposed by Barr, Levy, Scheepers and Tily (2013) was used. Per-child random adjustments to intercepts and slopes were included to account for individual differences. It was not possible to account for therapist effects due to the limited number of observations per therapist. When models did not converge, the number of iterations was increased and optimizer ‘Nelder_Mead’ was used instead of default optimizer ‘bobyqa’. *P* values were calculated with Type 3 sums of squares conditional F-tests and Kenward-Roger approximation for degrees of freedom using the function mixed from the package afex (version 0.20.2; Singmann, Bolker, Westfall, & Aust, 2018) which calls the function KRmodcomp of the package pbkrtest (version 0.4.7; Halekoh & Højsgaard, 2014). All models were run for four RQA measures, therefore, an alpha level of 0.05/4 = 0.0125 was used to account for multiple testing. Significant interaction effects were further explored using post-hoc pairwise comparisons provided by the packages emmeans (version1.2.1; Lenth, 2018), with Kenward-Roger approximation for degrees of freedom and Tukey *p* value correction.

Results

Baseline group characteristics

Children in the control group were slightly younger (M_{age} = 9.17, SD_{age} = 1.16) than children in the treatment group (M_{age} = 9.83, SD_{age} = 1.56), t (73.79) = -2.17, p = 0.034. Sex was distributed equally (treatment group 51% female, control group 56% female) among the groups, Chi-square (1) = 0.05, p = 0.825. Eighteen therapist delivered, on average, 9.10 (SD = 3.25) sessions per child. Children provided 654 samples of speech (control group 78, treatment group 576), over 338 conversations or therapy sessions (control group 41, treatment group 297).

Children in the treatment group had higher pre-treatment personalized anxiety scores (highest standardized SCARED-NL subscale; M = 1.39, SD = 0.90) than children in the control group (M = 0.32, SD = 1.01), t (77.71) = -4.97, p < .001.

Descriptives

Means, standard deviations and number of observations for variables that were included in the analyses can be found in Table 1, 2 and 3. Due to non-convergence of the RQA-parameterization, one amplitude time series and six frequency time series could not be analyzed using RQA, and were excluded from the analyses. Pairwise correlations between all RQA measures indicated that correlations within amplitude RQA measures and within frequency RQA measures were strong to very strong related. Correlations across voice characteristics were weaker, and frequency V_entr or frequency L_entr were least related to other RQA measures (see Table 4).

Table 1

Descriptives of RQA measures included in the analysis regarding anxiety levels and voice characteristics (first individual session for treatment group)

	Control group				Treatment group			
	Fear-related		Neutral		Fear-related		Neutral	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Amplitude DET	37	0.55 (0.17)	41	0.55 (0.16)	39	0.64 (0.16)	37	0.70 (0.13)
Amplitude LAM	37	0.70 (0.12)	41	0.69 (0.12)	39	0.76 (0.11)	37	0.80 (0.09)
Amplitude L_entr	37	0.94 (0.30)	41	0.92 (0.26)	39	1.13 (0.30)	37	1.26 (0.26)
Amplitude V_entr	37	1.45 (0.36)	41	1.43 (0.34)	39	1.69 (0.36)	37	1.79 (0.27)
Frequency DET	36	0.56 (0.14)	41	0.59 (0.14)	38	0.64 (0.12)	36	0.62 (0.13)
Frequency LAM	36	0.67 (0.13)	41	0.70 (0.11)	38	0.74 (0.10)	36	0.72 (0.12)
Frequency L_entr	36	1.21 (0.22)	41	1.23 (0.21)	38	1.24 (0.20)	36	1.28 (0.25)
Frequency V_entr	36	1.31 (0.33)	41	1.37 (0.28)	38	1.51 (0.25)	36	1.45 (0.33)

Table 2

Descriptives of RQA measures included in the analysis within the treatment group (all therapy sessions are included)

	Improvers				Non-improvers			
	Fear-related		Neutral		Fear-related		Neutral	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Amplitude DET	168	0.70 (0.15)	162	0.70 (0.15)	61	0.58 (0.16)	57	0.62 (0.15)
Amplitude LAM	168	0.80 (0.10)	162	0.80 (0.11)	61	0.72 (0.12)	57	0.75 (0.11)
Amplitude L_entr	168	1.24 (0.29)	162	1.25 (0.30)	61	1.03 (0.28)	57	1.09 (0.26)

(continued)

	Improvers				Non-improvers			
	Fear-related		Neutral		Fear-related		Neutral	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Amplitude V_entr	168	1.81 (0.35)	162	1.79 (0.34)	61	1.58 (0.35)	57	1.63 (0.34)
Frequency DET	166	0.61 (0.12)	162	0.62 (0.13)	61	0.65 (0.15)	56	0.61 (0.14)
Frequency LAM	166	0.73 (0.09)	162	0.74 (0.11)	61	0.76 (0.12)	56	0.72 (0.11)
Frequency L_entr	166	1.20 (0.22)	162	1.23 (0.26)	61	1.24 (0.23)	56	1.22 (0.24)
Frequency V_entr	166	1.50 (0.25)	162	1.49 (0.28)	61	1.59 (0.29)	56	1.49 (0.26)

Note. Data in this table is not independent. Twenty-seven children (20 improved, 7 non-improved) provided segments in multiple sessions. Time effects are neglected in this table.

Table 3

Descriptives of RQA measures included in the analysis comparing the first and last individual therapy sessions

	Fear-related segments							
	Improvers				Non-improvers			
	First session		Last session		First session		Last session	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Amplitude DET	20	0.66 (0.15)	20	0.75 (0.13)	6	0.60 (0.14)	7	0.60 (0.21)
Amplitude LAM	20	0.77 (0.11)	20	0.83 (0.08)	6	0.74 (0.09)	7	0.73 (0.15)
Amplitude L_entr	20	1.18 (0.30)	20	1.37 (0.27)	6	1.03 (0.31)	7	1.02 (0.29)
Amplitude V_entr	20	1.75 (0.36)	20	1.93 (0.29)	6	1.62 (0.41)	7	1.60 (0.39)
Frequency DET	19	0.62 (0.15)	20	0.63 (0.09)	6	0.68 (0.05)	7	0.67 (0.14)
Frequency LAM	19	0.74 (0.12)	20	0.75 (0.07)	6	0.76 (0.03)	7	0.77 (0.11)
Frequency L_entr	19	1.21 (0.23)	20	1.20 (0.15)	6	1.30 (0.07)	7	1.27 (0.17)
Frequency V_entr	19	1.54 (0.28)	20	1.51 (0.18)	6	1.49 (0.31)	7	1.60 (0.29)

	Neutral segments							
	Improvers				Non-improvers			
	First session		Last session		First session		Last session	
	n	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*	*n*	*M (SD)*
Amplitude DET	19	0.74 (0.11)	18	0.69 (0.14)	6	0.64 (0.13)	7	0.58 (0.13)
Amplitude LAM	19	0.83 (0.07)	18	0.79 (0.10)	6	0.76 (0.09)	7	0.72 (0.09)
Amplitude L_entr	19	1.33 (0.26)	18	1.21 (0.30)	6	1.13 (0.25)	7	1.04 (0.19)
Amplitude V_entr	19	1.87 (0.24)	18	1.76 (0.35)	6	1.71 (0.27)	7	1.55 (0.24)
Frequency DET	19	0.61 (0.12)	18	0.62 (0.14)	5	0.58 (0.10)	7	0.56 (0.13)
Frequency LAM	19	0.72 (0.14)	18	0.75 (0.10)	5	0.66 (0.10)	7	0.68 (0.10)
Frequency L_entr	19	1.27 (0.26)	18	1.23 (0.23)	5	1.16 (0.08)	7	1.16 (0.21)
Frequency V_entr	19	1.44 (0.38)	18	1.53 (0.27)	5	1.39 (0.31)	7	1.32 (0.19)

Table 4

Pairwise Pearson correlations for clustered data between RQA measures

	Amplitude				Frequency			
	DET	LAM	L_entr	V_entr	DET	LAM	L_entr	V_entr
Amplitude								
DET								
LAM	.99							
L_entr	.95	.93						
V_entr	.97	.95	.97					
Frequency								
DET	.33	.33	.28	.28				
LAM	.32	.31	.27	.28	.95			
L_entr	.20	.20	.15	.15	.79	.67		
V_entr	.12	.10	.13	.12	.78	.80	.49	

Note. n = 647. Observations of the RQA measures are not independent, because children provided multiple data points (two types of segments and several therapy sessions). Therefore, correlation coefficients for clustered data were calculated with the function pearson.clust that is meant for clustered data (Lorenz, Datta, & Harkema, 2011). *P* values are not provided within this approach.

Anxiety levels and voice characteristics

It was analyzed whether voice characteristics of children in the control group differed from voice characteristics of children in the treatment group during the first individual session. For each RQA measure (DET, LAM, L_entr, V_entr) for each voice characteristic (amplitude, frequency) a separated model was fitted, yielding eight models. Fixed effects of age (acting as a control variable), group (treatment, control) and segment type (fear-related, neutral), and their interactions were included. For each child, a random intercept was included. Models with amplitude had 154 observations from 80 children, models with frequency had 151 observations from 80 children. Model diagnostics of all models were inspected and gave no reason for concern.

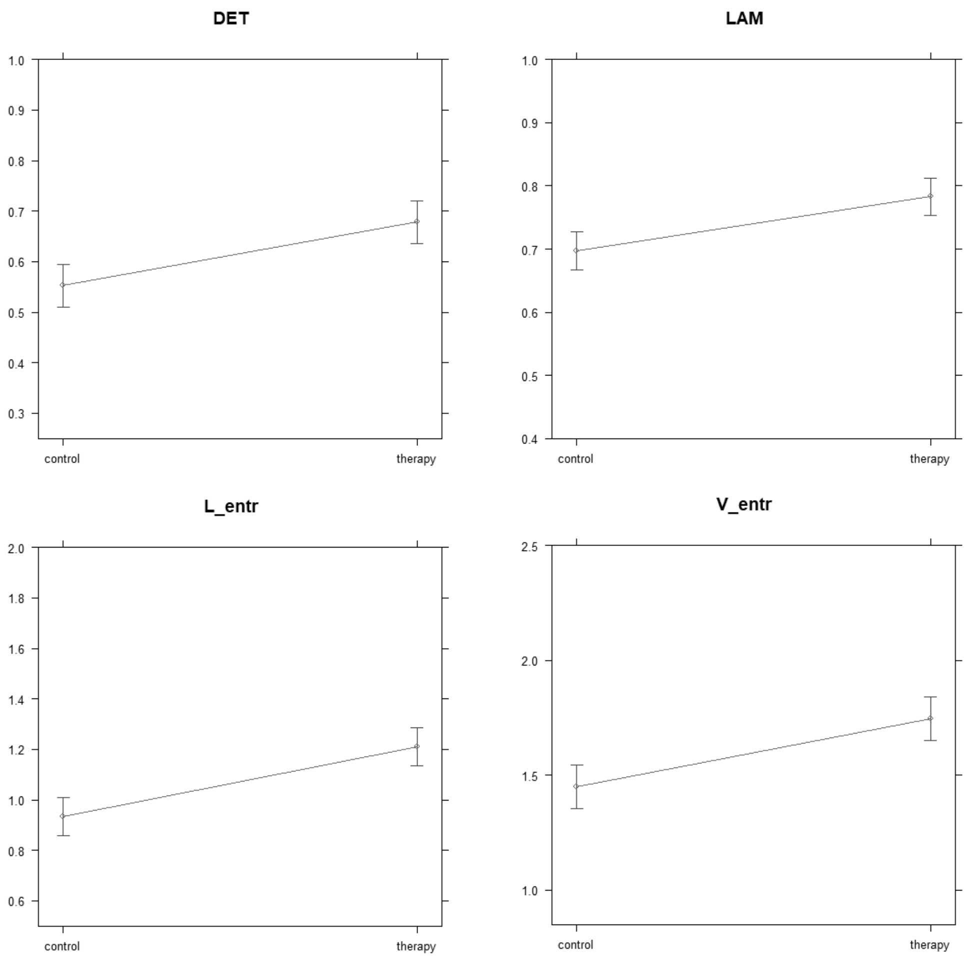

Figure 3. Model-based means with confidence intervals for RQA measures, effect of group.

Regarding amplitude, children in the treatment group had significantly higher DET (F(1, 76.88) = 17.21, p < .001), LAM (F(1,75.85) = 15.86, p <.001), L_entr (F(1,75.89) = 26.34, p < .001) and V_entr (F(1, 75.99) = 19.40, p <.001), than control children (see Figure 3). Regarding frequency, children in the treatment group did not differ in DET (F(1,75.11) = 3.17, p = .079), LAM (F(1, 75.19) = 2.60, p = .111), L_entr (F(1, 74.39) = 0.41, p = .522), or V_entr (F(1,74.86) = 4.56, p = .036). The interaction between segment type and group regarding amplitude L_entr (F(1,73.41) = 7.60, p = .007) was significant, all other p > .0125 [2]. Post-hoc tests indicated the amplitude L_entr of neutral segments from the treatment group

[2] Complete output (estimates and tests) for all models can be found in Appendix B.

(M = 1.28, SD = 0.05) was significantly higher than neutral (M = 0.91, SD = 0.04, t(130.19) = 5.83, p < .001) and fearful segments (M = 0.97, SD = 0.05, t(133.14) = 4.77, p <.001) from the control group. Amplitude L_entr of fearful segments from the treatment group (M = 1.13, SD = 0.05) was significantly higher than neutral segments from the control group, t(128.42) = 3.69, p = 0.002 (see Figure 4).

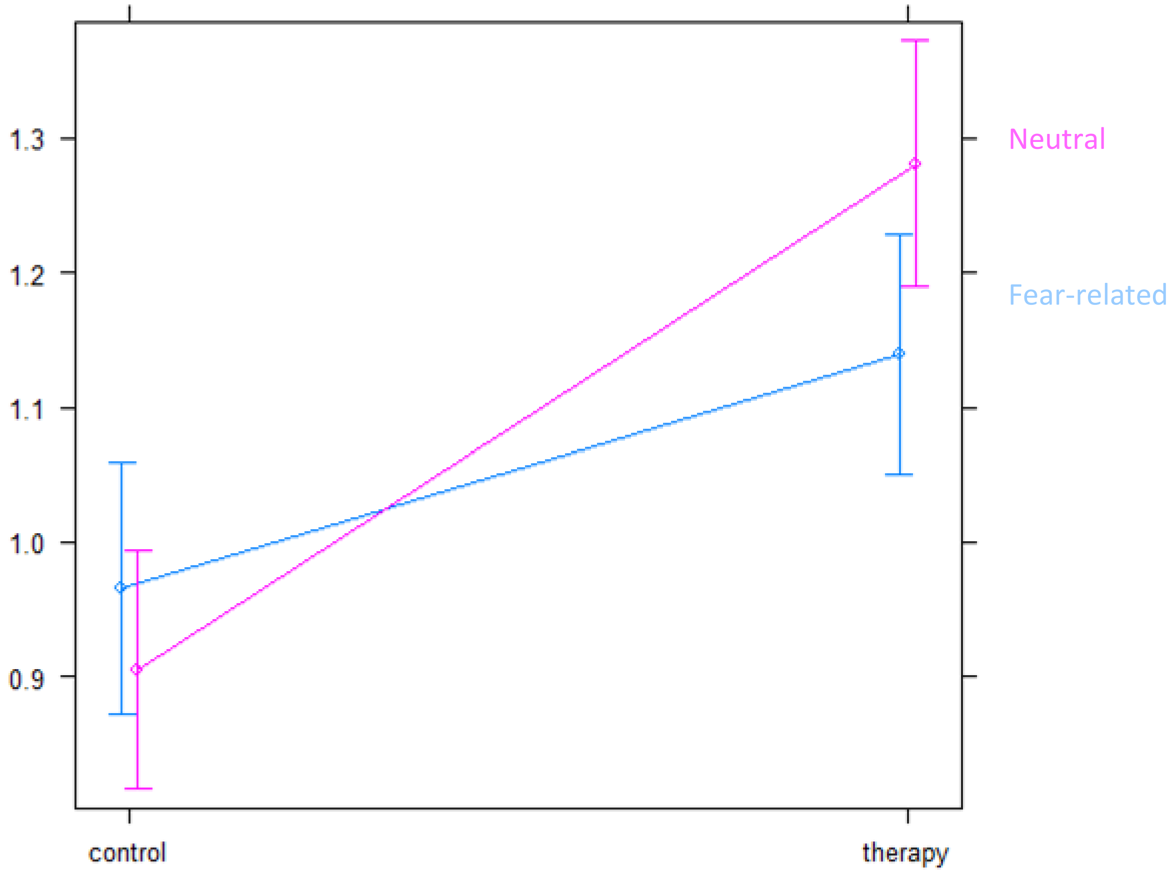

Figure 4. Model-based means with confidence intervals for amplitude L_entr, interaction effect between group and segment type.

Change over time

Within the treatment group, differences over time were analyzed. Mixed-effects models were fitted for each of the eight dependent variables. Fixed effects of age (acting as a control variable), segment type (fear-related, neutral), session number, and improver status (yes, no) and the interactions between segment type, session number, and improver status were included as predictors. Random intercepts for children and random effects of both session number and segment type varying over children were included. Models with amplitude had 448 observations from 27 children, models with frequency had 445 observations from 27 children. Upon inspection of model diagnostics, 2 influential and outlying L_entr frequency observations were excluded. All other model diagnostics gave no reason for concern.

Regarding amplitude, children who improved during treatment had higher DET ($F(1, 23.29) = 8.13$, $p = .009$) and L_entr ($F(1, 23.41) = 8.03$, $p = .009$), but not LAM ($F(1, 23.21) = 7.01$, $p = 0.014$) and V_entr ($F(1, 23.29) = 6.46$, $p = .018$) than children who did not improve (see Figure 5). Regarding frequency, improver status did not have significant effects on DET ($F(1, 22.83) = 0.54$, $p = .469$), LAM ($F(1, 23.09) = 0.06$, $p = .816$), L_entr ($F(1, 22.95) = 0.43$, $p = .520$), and V_entr ($F(1, 22.75) = 0.45$, $p = .511$).

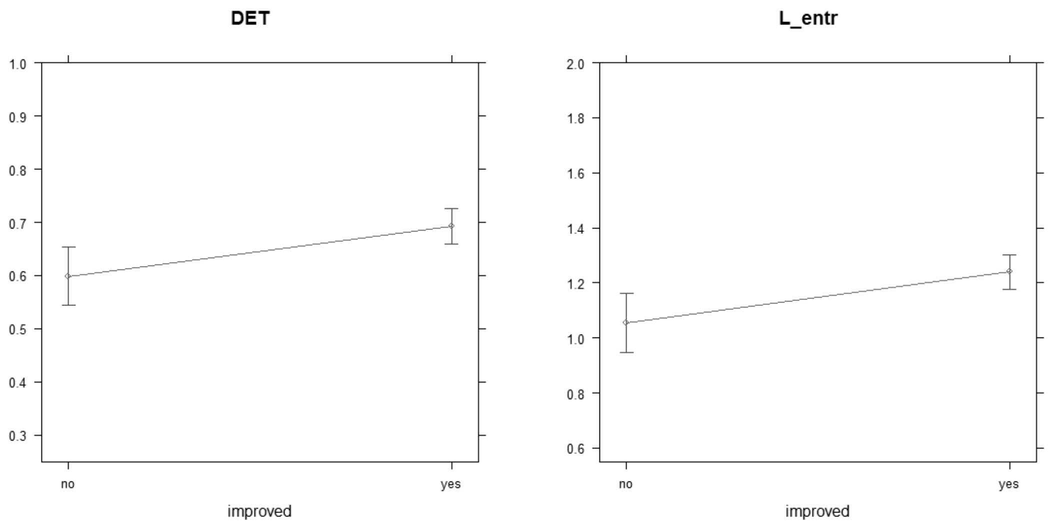

Figure 5. Model-based means with confidence intervals for amplitude DET and L_entr, effect of improver status.

To get better understanding of time effects, differences between the first and last individual sessions were analyzed. Fixed effects of age (acting as a control variable), segment type (fear-related, neutral), session (first, last), and improver status (yes, no) and the interactions between segment type, session, and improver status were included as predictors.. Random intercepts for child were included. Models with amplitude had 103 observations from 27 children, models with frequency had 101 observations from 27 children. Optimizer 'Nelder_Mead' had to be used for frequency L_entr in order to converge. Model diagnostics of all models were inspected and gave no reason for concern. However, none of the main effects or interactions was significant in any of the models.

Discussion

The main aim of the present study was to find RQA measures of voice characteristics that are indicative of anxiety levels, using audio segments from therapy sessions with anxious children and additional audio segments from conversations with control children. Amplitude but not frequency RQA measures of anxious children were higher than those of control children. Only one interaction effect was found, indicating that amplitude L_entr was highest for neutral segments of the treatment group. Within the treatment group, amplitude DET and L_entr of improvers were higher than those of non-improvers. In contrast to our expectations, voice characteristics did not change over time.

Taken together, these results indicate that RQA measures of voice amplitude differed between groups (treatment and control; improvers and non-improvers). However, both children in the control group and children who did not improve due to treatment had lower RQA measures. Given that improvers' anxiety levels are lower at the end of the treatment, we had expected that they would become more similar to children in the control group. Importantly, RQA measures did not change over time, nor were they sensitive to improvement due to treatment. Zooming in on the specific amplitude RQA measures that differed significantly, higher DET means that sequences in the signal were repeated and higher L_entr means that the length of these sequences was very irregular. Higher LAM means that the same value in the signal is maintained, whereas higher V_entr means that the length of these sequences of the same value is very irregular.

Our findings regarding frequency are not in line with previous research in which frequency RQA measures were predictive of several other types of psychopathology (Fusaroli, et al., 2013). Three differences between their study and our approach should be noted. First, Fusaroli and colleagues used a machine learning approach with principal component analysis that included many voice features. In the current paper, we focused on RQA measures, because these might be particularly suited to analyze voice features related to anxiety or improvement due to treatment. Second, Fusaroli et al. used a standardized laboratory setting (i.e. people freely described videos with moving shapes) which may have facilitated the detection of differences between groups of people. Although we wanted to investigate voices in a naturalistic setting, this might have precluded us from finding effects. To understand the relationship between voice characteristics and anxiety, it might be necessary to first use more controlled settings. Third, whereas Fusaroli et al. used mean fundamental frequency we chose to analyze instantaneous frequency that emphasizes local characteristics of the signal. Advantages of this instantaneous frequency are that the underlying signal is described much more detailed regarding frequencies present in the data and that it was developed to be used for complex signals, such as voices. However, our results point towards the fact that instantaneous frequency may not be able to capture voice characteristics appropriately. To the contrary, mean fundamental frequency was successfully used by Fusaroli et al. and this measure is known to be closely related to the pitch that humans can hear in voices, making interpretation more straightforward.

That we did not detect pitch-related differences between control children and anxious children is consistent with Scharstein et al. (2011) and Scharstein and Beidel (2014) who did not find differences in pitch between control children and, respectively, children with social anxiety and children with generalized anxiety disorder. Yet, the later study also included children with social anxiety and these did differ from control children. Several other studies indicated relationships between anxiety levels and pitch in socially anxious adults as well (Laukka, et al., 2008; Weeks, et al. 2012, Weeks et al., 2016). This may give rise to the hypothesis that exclusively social anxiety is related to pitch differences, which we could not check due to a limited sample size. In addition, effects in previous studies often concerned the range of pitch that people employ, with higher anxiety associated with more restricted pitch ranges. Because we scaled the pitch of each segment to range between 0 and 1, we could not assess whether this was true for our sample. Also, RQA measures reflect the structure within the signal, not the absolute values that are present.

Our findings regarding amplitude (loudness) corroborate with results from an archived bachelor thesis (Kroytor, 2012). However, in that study, anxious children's voices became slightly more similar to the control group during treatment. Lastly, it is difficult to assess to what degree our results are congruent with Flykt et al. (2017) who found differences regarding amplitude in response to pictures of feared compared to non-feared animals in people with specific phobia. We did not find main effects of fear-related versus neutral conversation topics, and only one interaction effect involving the topic turned significant. In addition, their experimental setup confronted people with sudden images of feared animals upon which they to react as fast as possible, whereas talking about a fear-related situation in a therapy session is much more deliberate.

In light of earlier research on rigidity related to anxiety, our findings are difficult to understand. The combination of high determinism and high entropy in anxious children's voices indicates that there is high regularity but at the same time a very unpredictable signal. This is in contradiction with findings regarding heart rate, where anxiety was related to lower variability and complexity (i.e. more predictable) heart rate (Pittig, et al., 2013; Bornas, et al., 2015). Voice and heart rate complexity may have different relations with anxiety.

The current study has several strengths. An important aspect is the data collection method. Audio recording is a non-invasive, easy and cheap method to gather data. It captured real behavior and did not interfere with treatment. Compared to the video describing task used by Fusaroli et al. (2013), audiotaped therapy fragments are much more naturalistic and unobtrusive. Children talked about wide-ranging topics, according to their individual fears, making the assessment an individualized measure. We also followed children over time, which increased the power of our design due to the within-person data and made it possible to see treatment effects. We extracted multiple audio segments within each child, compared anxious children to typically developing children, and looked at changes in voice characteristics across treatment. Given that self-reported anxiety levels and physiological measures often do not match in anxiety disorders (Pittig, et al., 2013), it is important to include other predictors than merely self-reported questionnaires. In our study, we looked at two different groups (treatment and control) and two types of segments (fear-related and neutral) in addition to self-reported anxiety.

We are aware that our study may have several limitations. Firstly, only half of the available data from the treatment group could be included until now, reducing power to detect effects. Secondly, our definition of improver status was not optimal. We used self-reported anxiety measurement to define whether children improved or not, because diagnostic interviews were not included in the project from which the treatment data came (van Doorn, 2017). Future research should include clinical interviews or therapist reports to achieve more reliable outcome measures. Thirdly, we did not collect data from the control group over time so we could not compare developments in both groups. Including multiple measurement moments for control children would, for example, make it possible to compare stability between control and treatment group. Fourthly, conversation topics varied a lot. Fear-related segments could cover the worst fear of the child, or something quite unimportant, whereas neutral segments could cover a wide range of neutral to positive topics. It would be good to include several segments of each type per session, to see whether the voice characteristics are be more stable within than between sessions. Once again, this demonstrates the need for research in a more controlled setting than the naturalistic situation we used.

Taking these limitations into account, the current study was the first to identify differences in vocal RQA measures in relation to anxiety levels. However, the RQA measures were not sensitive to changes in anxiety due to treatment, which was the most important aim of this study. In addition, quantifying pitch-related features of voices with the use of instantaneous frequency does not seem to work: mean fundamental frequency may be more suited. Still, our findings indicate that it may be possible to use voice characteristics as unobtrusive measurements of anxiety. However, testing this in a naturalistic environment firstly may have been a step too far. To know better whether and how RQA voice characteristics relate to anxiety, research in a more controlled setting is needed. For example, children could be asked to describe their worst fears. Although this would mean that children need to complete additional tasks, these manipulations may be necessary to understand voice dynamics.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Analysis of Factorial Experiments. R package version 0.20-2.

Bettis, A. H., Forehand, R., Sterba, S. K., Preacher, K. J., & Compas, B. E. (2016). Anxiety and depression in children of depressed parents: Dynamics of change in a preventive intervention. *Journal of Clinical Child & Adolescent Psychology*, 1-14. https://doi.org/10.1080/15374416.2016.1225503

Birmaher, B., Brent, D. A., Chiappetta, L., Bridge, J., Monga, S., & Baugher, M. (1999). Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. J*ournal of the American Academy of Child & Adolescent Psychiatry*, 38(10), 1230-1236.

Bögels, S. M. (2008). *Behandeling van angststoornissen bij kinderen en adolescenten. Met het cognitief gedragstherapeutisch protocol Denken + Doen = Durven* [Treatment of anxiety disorders among children and adolescents. Using the cognitive-behavioural therapeutic manual Thinking + Doing = Daring]. Houten: Bohn Stafleu van Loghum.

Bornas, X., Balle, M., De la Torre-Luque, A., Fiol-Veny, A., & Llabres, J. (2015). Ecological assessment of heart rate complexity: Differences between high- and low-anxious adolescents. *International Journal of Psychophysiology*, *98*, 112-118.

Bugental, D. B., Beaulieu, D. A., Schwartz, A., & Dragosits, R. (2009). Domain-specific responses to power-based interaction. *Journal of Experimental Social Psychology*, *45*, 386–391. https://doi.org/10.1016/j.jesp.2008.10.005

Copeland, W. E., Adair, C. E., Smetanin, P., Stiff, D., Briante, C., Colman, I., … & Angold, A. (2013). Diagnostic transitions from childhood to adolescence to early adulthood. Journal of Child Psychology and Psychiatry, *54*, 791–9.

Donghoh, K. & Hee-Seok, O. (2014). EMD: Empirical Mode Decomposition and Hilbert Spectral Analysis. R package version 1.5.7.

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, *4*, 973.

Flykt, A., Bänziger, T., & Lindeberg, S. (2017). Intensity of vocal responses to spider and snake pictures in fearful individuals. *Australian Journal of Psychology*, *69*, 184-191.

Friedman, B. H., & Thayer, J. F. (1998). Anxiety and autonomic flexibility: a cardiovascular approach1. *Biological psychology*, *47*, 243-263.

Fusaroli, R., Grossman, R., Cantio, C., Bilenberg, N., & Weed, E. (2015, May). *The temporal structure of the autistic voice: A cross-linguistic investigation.* Poster session presented at the International Meeting for Autism Research, Salt Lake City.

Fusaroli, R., Tylèn, K., Simonsen, A., & Weed, E. (2013, August). *Non-linear dynamics of voice in mental disorders*. Poster session presented at the Annual Meeting of the Cognitive Science Society, Berlin.

Galili, L., Amir, O., & Gilboa-Schechtman, E. (2013). Acoustic properties of dominance and request utterances in social anxiety. *Journal of Social and Clinical Psychology*, *32*, 651-673.

Gilboa-Schechtman, E., Galili, L., Sahar, Y., & Amir, O. (2014). Being “in” or “out” of the game: subjective and acoustic reactions to exclusion and popularity in social anxiety. *Frontiers in Human Neuroscience*, *8*, 147.

Hasselman, F. (2018). Casnet: A Toolbox for Studying Complex Adaptive Systems and NETworks.

Imel, Z. E., Barco, J. S., Brown, H. J., Baucom, B. R., Baer, J. S., Kircher, J. C., & Atkins, D. C. (2014). The association of therapist empathy and synchrony in vocally encoded arousal. *Journal of Counseling Psychology, 61*, 146-153.

Jackson, E. S., Tiede, M., Beal, D., & Whalen, D. H. (2016). The impact of social–cognitive stress on speech variability, determinism, and stability in adults who do and do not stutter. *Journal of Speech, Language, and Hearing Research*, *59*, 1295-1314.

James, A. C., James, G., Cowdrey, F. A., Soler, A., & Choke, A. (2015). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane database of systematic reviews, 2,* 1-87.

Jansen, M., van Doorn, M. M., Lichtwarck-Aschoff, A., Kuijpers, R. C., Theunissen, H., Korte, M., ... & Granic, I. (2012). Effectiveness of a cognitive-behavioral therapy (CBT) manualized program for clinically anxious children: study protocol of a randomized controlled trial. *BMC psychiatry*, *12*, 16.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H. U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International journal of methods in psychiatric research*, 21, 169-184.

Kroytor, A. (2012). *Does behavioral treatment for children with social anxiety disorder change vocal characteristics?* (Bachelor’s thesis). Retrieved from http://stars.library.ucf.edu/honorstheses1990-2015/1776.

Laukka, P., Linnman, C., Åhs, F., Pissiota, A., Frans, Ö., Faria, V., ... & Furmark, T. (2008). In a nervous voice: Acoustic analysis and perception of anxiety in social phobics’ speech. *Journal of Nonverbal Behavior, 32,* 195.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., Moroz, I.M. (2007). Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *BioMedical Engineering OnLine*, *6*, 23.

Lopes, L. W., Costa, S. L. D. N. C., de Almeida Costa, W. C., Correia, S. É. N., & Vieira, V. J. D. (2014). Acoustic assessment of the voices of children using nonlinear analysis: proposal for assessment and vocal monitoring. *Journal of Voice*, *28,* 565-573.

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics reports, 438*, 237-329.

Pittig, A., Arch, J. J., Lam, C. W., & Craske, M. G. (2013). Heart rate and heart rate variability in panic, social anxiety, obsessive–compulsive, and generalized anxiety disorders at baseline and in response to relaxation and hyperventilation. International *Journal of Psychophysiology, 87*, 19-27.

R core team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rufiner, H. L., Torres, M. E., Gamero, L., & Milone, D. H. (2004). Introducing complexity measures in nonlinear physiological signals: application to robust speech recognition. *Physica A: Statistical Mechanics and its Applications, 332*, 496-508.

Russell Lenth (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.2.1. https://CRAN.R-project.org/package=emmeans

Safford, S. M., Kendall, P. C., Flannery-Schroeder, E., Webb, A., & Sommer, H. (2005). A longitudinal look at parent–child diagnostic agreement in youth treated for anxiety disorders. *Journal of Clinical Child and Adolescent Psychology, 34*, 747-757.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American psychologist, 54,* 93.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4,* 1-32.

signal developers (2013). signal: Signal processing. http://r-forge.r-project.org/projects/signal/.

Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). afex: Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical systems and turbulence*, Warwick 1980 (pp. 366-381). Springer, Berlin, Heidelberg.

Turvey, M. T. (2007). Action and perception at the level of synergies. *Human movement science, 26*, 657-697.

Halekoh, U. & Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest., *Journal of Statistical Software, 58*, 1-30. http://www.jstatsoft.org/v59/i09/

Ligges, U., Krey, S., Mersmann, U. & Schnackenberg, S. (2016). tuneR: Analysis of music. http://r-forge.r-project.org/projects/tuner/.

van Doorn, M. E. M., Jansen, M., Bodden, D. H. M., Lichtwarck-Aschoff, A., & Granic, I. (2017). A randomized controlled effectiveness study comparing manualized cognitive behavioral therapy (CBT) with treatment-as-usual for clinically anxious children. *Journal of Clinical Trials, 7*.

Weeks, J. W., Lee, C. Y., Reilly, A. R., Howell, A. N., France, C., Kowalsky, J. M., & Bush, A. (2012). “The Sound of Fear”: Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder. *Journal of Anxiety Disorders, 26,* 811-822.

Weeks, J. W., Srivastav, A., Howell, A. N., & Menatti, A. R. (2016). “Speaking More than Words”: Classifying Men with Social Anxiety Disorder via Vocal Acoustic Analyses of

Diagnostic Interviews. *Journal of Psychopathology and Behavioral Assessment, 38*, 30-41.

Wenze, S. J., & Miller, I. W. (2010). Use of ecological momentary assessment in mood disorders research. *Clinical psychology review, 30*, 794-804.

Zbilut, J. P., & Webber Jr, C. L. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics letters A*, *171*, 199-203.

Appendix A

```
# Packages
library(rio)
library(tuneR)
library(EMD)
library(plyr)
library(wmtsa)
library(casnet)
library(signal)

s1<- readMP3(s)

## Highpass filter
fltr   <- function(f,x){signal::filtfilt(filt=f, x=x)}
fs <- 44100 # sample rate
fc <- 100; # cut-off frequency
N  <- 20; # order
hp <- butter(N,1/(1+(fs/fc)^(2*N)),"high") # calculation of coeffcientes for Butterworth filter.
s1[[2]]@left <- fltr(f=hp,x=na.exclude(s1[[2]]@left))

## Downsample audio
s1d <- downsample(s[[2]],samp.rate=2000)

## Maximal length
if (length(s1d[[2]]@left) > 10000) { s1d[[2]]@left <- s1d[[2]]@left[1:10000]

## Hilbert transform (analytic signal: amplitude and instantaneous frequency)
tt <- cbind((1:length(s1d[[2]]@left))*(1/s1d[[2]]@samp.rate))
y  <- hilbertspec(xt=cbind(s1d[[2]]@left),tt=tt
ts_ampl <- list(s1d[[1]], ts(data=y$amplitude,start=0,frequency=s1d[[2]]@samp.rate))
ts_instfreq <- list(s1d[[1]], ts(data=y$instantfreq,start=0,frequency=s1d[[2]]@samp.rate
ts_y<- list(s1d[[1]],
ts(data=abs(y$amplitude)/max(abs(y$amplitude)),start=0,frequency=s1d[[2]]@samp.rate))
ts_yfreq<- list(s1d[[1]],
ts(data=abs(y$instantfreq)/max(abs(y$instantfreq)),start=0,frequency=s1d[[2]]@samp.rate))

#Optimal delay (using nonlinearTseries): first minimum, global minimum, maximal lag
ts_y_with_lag<-est_emLag(ts_y[[2]])

#Optimal dimensions (using nonlinearTseries)
ts_y_with_dim<-est_emDim(ts_y_with_lag[[2]],delay = 4,do.plot=F)

#RQA using command line
crqa_output<-
crqa_cl(as.numeric(ts_y_with_dim[[2]]),emDim=ts_y_with_dim[[4]],emLag=4,emRad=NA,tar
getValue=0.01)
```

Appendix B

Below, estimated parameters and their standard errors are shown for the mixed-effects models, including conditional F-tests with Kenward-Roger approximation of degrees of freedom.

Table A1. Anxiety levels and voice characteristics, regarding amplitude

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.62	0.02				
Age (centered)	0.01	0.01	1	75.54	0.38	.540
Group	-0.06	0.02	1	76.88	17.21	< .001
Segment type	-0.01	0.01	1	73.42	0.82	.367
Age x Group	0.01	0.01	1	75.54	1.61	.209
Age x Segment type	0.01	0.01	1	73.12	3.44	.068
Group x Segment type	0.02	0.01	1	73.42	4.50	.037
Age x Group x Segment type	0.01	0.01	1	73.12	1.49	.227
	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.74	0.01				
Age (centered)	-0.00	0.01	1	75.50	0.39	.535
Group	-0.04	0.01	1	75.85	15.86	<.001
Segment type	-0.01	0.01	1	73.51	0.82	.369
Age x Group	0.01	0.01	1	75.50	1.58	.212
Age x Segment type	0.01	0.01	1	73.20	2.99	.088
Group x Segment type	0.01	0.01	1	73.51	3.79	.055
Age x Group x Segment type	0.01	0.01	1	73.20	1.28	.261

	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.07	0.02				
Age (centered)	-0.01	0.02	1	75.55	0.65	.424
Group	-0.13	0.03	1	75.89	26.34	<.001
Segment type	-0.02	0.02	1	73.41	1.23	.271
Age x Group	0.02	0.02	1	75.55	1.61	.208
Age x Segment type	0.03	0.01	1	73.11	6.23	.015
Group x Segment type	0.05	0.01	1	73.41	7.60	.007
Age x Group x Segment type	0.01	0.01	1	73.11	0.48	.490
	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.60	0.03				
Age (centered)	-0.00	0.02	1	75.68	.03	.873
Group	-0.15	0.03	1	75.99	19.40	<.001
Segment type	-0.01	0.02	1	73.08	0.45	.504
Age x Group	0.02	0.02	1	75.68	0.52	.471
Age x Segment type	0.03	0.02	1	72.80	4.19	.044
Group x Segment type	0.04	0.02	1	73.08	3.99	.050
Age x Group x Segment type	0.02	0.02	1	72.80	1.29	.260

Table A2. Anxiety levels and voice characteristics, regarding frequency

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.61	0.01				
Age (centered)	0.01	0.01	1	74.74	1.14	.289
Group	-0.02	0.01	1	75.11	3.17	.079
Segment type	0.00	0.01	1	72.15	0.04	.843
Age x Group	0.02	0.01	1	74.74	3.42	.068
Age x Segment type	0.00	0.01	1	71.80	0.45	.503
Group x Segment type	-0.01	0.01	1	72.15	2.41	.125
Age x Group x Segment type	0.00	0.01	1	71.80	0.19	.662
	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.71	0.01				
Age (centered)	0.01	0.01	1	74.82	1.22	.273
Group	-0.02	0.01	1	75.19	2.60	.111
Segment type	0.00	0.01	1	72.00	0.01	.919
Age x Group	0.01	0.01	1	74.83	2.21	.141
Age x Segment type	-0.01	0.01	1	71.66	0.75	.390
Group x Segment type	-0.01	0.01	1	72.00	3.21	.077
Age x Group x Segment type	0.00	0.01	1	71.66	0.04	.841
	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.25	0.02				
Age (centered)	0.03	0.01	1	74.00	2.64	.108

Group	-0.01	0.02	1	74.39	0.41	.522
Segment type	-0.01	0.02	1	73.24	0.35	.552
Age x Group	0.03	0.01	1	74.00	5.70	.020
Age x Segment type	-0.02	0.01	1	72.85	2.04	.157
Group x Segment type	0.00	0.02	1	73.24	0.00	.979
Age x Group x Segment type	0.00	0.01	1	72.85	0.07	.798

	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.42	0.03				
Age (centered)	0.04	0.02	1	74.48	3.80	.055
Group	-0.06	0.03	1	74.86	4.56	.036
Segment type	0.00	0.02	1	72.58	0.02	.878
Age x Group	0.03	0.02	1	74.48	1.69	.198
Age x Segment type	-0.03	0.02	1	72.21	3.07	.084
Group x Segment type	-0.04	0.02	1	72.58	3.27	.074
Age x Group x Segment type	0.01	0.02	1	72.21	0.08	.777

Table A3. Change over time, regarding amplitude

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.65	0.02				
Age (centered)	-0.01	0.01	1	23.04	0.90	.350
Segment type	-0.01	0.01	1	23.16	2.86	.104
Session (centered)	0.00	0.00	1	18.88	0.17	.684
Improver status	-0.05	0.02	1	23.30	8.13	.009
Segment type x Session	0.00	0.00	1	397.31	1.18	.278
Segment type x Improver status	-0.01	0.01	1	23.16	1.52	.230
Session x Improver status	0.00	0.00	1	18.88	0.07	.799
Segment type x Session x Improver status	0.00	0.00	1	397.11	0.35	.552
	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.76	0.01				
Age (centered)	0.01	0.01	1	23.13	0.76	.393
Segment type	-0.01	0.01	1	23.09	2.43	.133
Session (centered)	0.00	0.00	1	18.82	0.47	.500
Improver status	-0.03	0.01	1	23.21	7.01	.014
Segment type x Session	0.00	0.00	1	397.37	1.13	.288
Segment type x Improver status	-0.01	0.01	1	23.09	1.64	.213
Session x Improver status	0.00	0.00	1	18.82	0.13	.724
Segment type x Session x Improver status	0.00	0.00	1	397.24	0.35	.555

	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.148	0.03				
Age (centered)	-0.02	0.02	1	23.13	0.77	.390
Segment type	-0.02	0.02	1	23.12	1.90	.181
Session (centered)	0.00	0.00	1	18.76	0.15	.703
Improver status	-0.09	0.03	1	23.41	8.03	.009
Segment type x Session	0.01	0.00	1	398.08	1.27	.260
Segment type x Improver status	-0.01	0.02	1	23.12	0.84	.369
Session x Improver status	0.00	0.00	1	18.76	0.00	.950
Segment type x Session x Improver status	0.00	0.00	1	397.86	0.71	.400

	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.70	0.04				
Age (centered)	0.00	0.02	1	22.99	0.75	.395
Segment type	-0.01	0.02	1	23.18	0.50	.487
Session (centered)	0.00	0.01	1	18.85	0.20	.660
Improver status	-0.09	0.04	1	23.29	6.46	.018
Segment type x Session	0.01	0.01	1	397.69	1.10	.294
Segment type x Improver status	-0.02	0.02	1	23.18	1.10	.305
Session x Improver status	0.00	0.01	1	18.85	0.00	.994
Segment type x Session x Improver status	0.00	0.01	1	397.50	0.27	.602

Table A4. Change over time, regarding frequency

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.62	0.01				
Age (centered)	0.00	0.01	1	22.56	0.07	.793
Segment type	0.01	0.01	1	23.11	1.14	.297
Session (centered)	0.00	0.00	1	20.46	0.35	.561
Improver status	0.01	0.01	1	22.83	0.54	.469
Segment type x Session	0.00	0.00	1	394.392	0.72	.396
Segment type x Improver status	0.01	0.01	1	23.11	3.22	.086
Session x Improver status	0.00	0.00	1	20.44	1.68	.210
Segment type x Session x Improver status	0.00	0.00	1	394.25	0.08	.772
	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.74	0.01				
Age (centered)	0.00	0.00	1	21.46	0.06	.806
Segment type	0.01	0.01	1	23.19	1.29	.268
Session (centered)	0.00	0.00	1	21.77	0.00	.956
Improver status	0.00	0.01	1	23.09	0.06	.816
Segment type x Session	0.00	0.00	1	394.27	0.92	.338
Segment type x Improver status	0.01	0.01	1	23.19	3.55	.072
Session x Improver status	0.00	0.00	1	21.72	1.52	.230
Segment type x Session x Improver status	0.00	0.00	1	394.05	0.07	.798

	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.23	0.02				
Age (centered)	0.00	0.01	1	23.00	0.00	.980
Segment type	-0.01	0.01	1	22.84	0.33	.571
Session (centered)	0.00	0.00	1	20.10	1.43	.245
Improver status	0.01	0.02	1	22.95	0.43	.520
Segment type x Session	0.00	0.00	1	394.57	0.06	.802
Segment type x Improver status	0.01	0.01	1	22.84	1.15	.296
Session x Improver status	0.00	0.00	1	20.09	0.13	.721
Segment type x Session x Improver status	0.00	0.00	1	394.53	0.05	.828

	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.52	0.02				
Age (centered)	0.01	0.01	1	21.69	1.07	.313
Segment type	0.03	0.01	1	23.00	3.93	.060
Session (centered)	0.00	0.01	1	21.76	0.21	.657
Improver status	0.02	0.02	1	22.75	0.45	.511
Segment type x Session	0.00	0.00	1	394.58	0.00	.983
Segment type x Improver status	0.02	0.01	1	23.00	2.79	.109
Session x Improver status	0.00	0.01	1	21.71	0.20	.662
Segment type x Session x Improver status	0.00	0.00	1	394.54	0.61	.435

Table A5. Change over time (first and last session), regarding amplitude

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.66	0.02				
Age (centered)	-0.01	0.01	1	23.36	0.81	.378
Segment type	-0.01	0.02	1	24.99	0.19	.668
Session (first/last)	0.00	0.02	1	25.05	0.00	.970
Improver status	-0.05	0.02	1	25.31	5.26	.030
Segment type x Session	-0.03	0.01	1	23.88	5.99	.022
Segment type x Improver status	0.00	0.02	1	24.99	0.01	.910
Session x Improver status	0.01	0.02	1	25.02	0.39	.539
Segment type x Session x Improver status	0.01	0.01	1	23.88	0.62	.440
	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.77	0.01				
Age (centered)	-0.01	0.01	1	23.30	0.91	.351
Segment type	0.00	0.01	1	24.95	0.17	.685
Session (first/last)	0.00	0.01	1	25.10	0.04	.849
Improver status	-0.03	0.01	1	25.37	4.36	.047
Segment type x Session	-0.02	0.01	1	23.93	5.42	.029
Segment type x Improver status	0.00	0.01	1	24.96	0.00	.967
Session x Improver status	0.01	0.01	1	25.07	0.62	.440
Segment type x Session x Improver status	0.01	0.01	1	23.94	0.79	.382

	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.17	0.04				
Age (centered)	-0.03	0.02	1	23.40	2.36	.138
Segment type	-0.02	0.03	1	24.99	0.31	.582
Session (first/last)	0.00	0.03	1	24.88	0.02	.883
Improver status	-0.09	0.04	1	25.21	5.26	.030
Segment type x Session	-0.05	0.02	1	24.29	4.92	.036
Segment type x Improver status	-0.02	0.03	1	25.00	4.92	.036
Session x Improver status	0.02	0.03	1	24.86	0.40	.534
Segment type x Session x Improver status	0.03	0.02	1	24.29	1.08	.310

	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.72	0.05				
Age (centered)	-0.01	0.02	1	23.55	0.24	.630
Segment type	0.00	0.03	1	24.89	0.01	.939
Session (first/last)	0.01	0.03	1	24.62	0.05	.817
Improver status	-0.10	0.05	1	25.18	4.20	.051
Segment type x Session	-0.06	0.03	1	24.31	4.27	.050
Segment type x Improver status	-0.01	0.03	1	24.89	0.19	.665
Session x Improver status	0.03	0.03	1	24.60	0.56	.461
Segment type x Session x Improver status	0.01	0.03	1	24.30	0.27	.609

Table A6. Change over time (first and last session), regarding frequency

	DET					
	Estimate	SE	df1	df2	F	p
Intercept	0.62	0.02				
Age (centered)	-0.01	0.01	1	23.68	0.70	.412
Segment type	0.03	0.01	1	25.10	5.74	.024
Session (first/last)	0.00	0.01	1	24.69	0.00	.945
Improver status	0.00	0.02	1	25.16	0.05	.833
Segment type x Session	0.00	0.01	1	25.06	0.02	.882
Segment type x Improver status	0.02	0.01	1	25.09	3.45	.075
Session x Improver status	0.01	0.01	1	24.69	0.38	.544
Segment type x Session x Improver status	0.00	0.01	1	25.07	0.01	.940

	LAM					
	Estimate	SE	df1	df2	F	p
Intercept	0.73	0.01				
Age (centered)	-0.01	0.01	1	23.85	0.41	.530
Segment type	0.03	0.01	1	25.11	6.26	.019
Session (first/last)	-0.01	0.01	1	24.92	0.42	.522
Improver status	-0.01	0.02	1	25.50	0.16	.393
Segment type x Session	0.00	0.01	1	25.01	0.09	.769
Segment type x Improver status	0.02	0.01	1	25.11	3.53	.072
Session x Improver status	0.00	0.01	1	24.92	0.05	.825
Segment type x Session x Improver status	0.00	0.01	1	25.02	0.09	.761

	L_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.23	0.03				
Age (centered)	-001	0.02	1	23.69	0.59	.451
Segment type	0.02	0.02	1	25.31	0.78	.387
Session (first/last)	0.01	0.02	1	24.50	0.12	.733
Improver status	0.00	0.03	1	25.39	0.01	.912
Segment type x Session	0.00	0.02	1	24.93	0.02	.876
Segment type x Improver status	0.05	0.02	1	25.31	3.88	.060
Session x Improver status	0.00	0.02	1	24.51	0.02	.897
Segment type x Session x Improver status	0.00	0.02	1	24.94	0.03	.854

	V_entr					
	Estimate	SE	df1	df2	F	p
Intercept	1.49	0.04				
Age (centered)	-0.01	0.02	1	23.97	0.14	.716
Segment type	0.05	0.03	1	25.28	3.13	.089
Session (first/last)	-0.01	0.03	1	24.70	0.08	.786
Improver status	-0.02	0.04	1	25.85	0.19	.664
Segment type x Session	-0.01	0.03	1	25.00	0.10	.759
Segment type x Improver status	0.03	0.03	1	25.28	1.22	.280
Session x Improver status	0.01	0.03	1	24.70	0.08	.775
Segment type x Session x Improver status	-0.04	0.03	1	25.00	2.37	.136