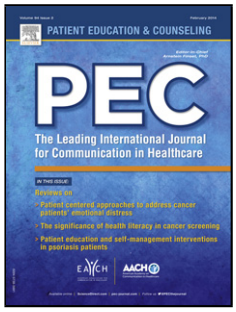


Title: Using fundamental frequency of cancer survivors' speech to investigate emotional distress in out-patient visits

Author: Jacqueline Kandsberger Simon N. Rogers Yuefang Zhou Gerry Humphris

PII: S0738-3991(16)30335-4
DOI: http://dx.doi.org/doi:10.1016/j.pec.2016.08.003
Reference: PEC 5421

To appear in: *Patient Education and Counseling*

Received date: 7-12-2015
Revised date: 15-7-2016
Accepted date: 2-8-2016

Please cite this article as: Kandsberger Jacqueline, Rogers Simon N, Zhou Yuefang, Humphris Gerry.Using fundamental frequency of cancer survivors' speech to investigate emotional distress in out-patient visits.*Patient Education and Counseling* http://dx.doi.org/10.1016/j.pec.2016.08.003

Title: Using fundamental frequency of cancer survivors' speech to investigate emotional distress in out-patient visits

Authors

Jacqueline Kandsberger; Medical School, University of St. Andrews KY16 9TF

Simon N. Rogers; Merseyside Regional Head & Neck Cancer Centre, Aintree Hospital, Liverpool, L9 7AL

Yuefang Zhou; Medical School, University of St Andrews KY16 9TF

Gerry Humphris; University of St Andrews, KY16 9TF and Edinburgh Cancer Centre, Western General Hospital EH4 2XU

Corresponding Author

Gerry Humphris

Medical School

University of St Andrews

North Haugh

St Andrews, KY16 9TF, UK

Tel: +44(0)1334463565

gmh4@st-andrews.ac.uk

Highlights

- Emotions in clinical conversations of cancer patients studied by a coding system and speech prosody
- Emotional energy f0 was associated with cues and concerns as coded by VR-CoDES system
- An additional aid to study emotional speech with potential to reveal hidden content and meaning

Abstract

Objective: Emotions, are in part conveyed by varying levels of fundamental frequency of voice pitch (*f0*). This study tests the hypothesis that patients display heightened levels of emotional arousal (*f0*) during Verona Coding Definitions of Emotional Sequences (VR-CoDES) cues and concerns versus during neutral statements.

Methods: The audio recordings of sixteen head and neck cancer survivors' follow-up consultations were coded for patients' emotional distress. Pitch (*f0*) of coded cues and concerns, including neutral statements was extracted. These were compared using a hierarchical linear model, nested for patient and pitch range, controlling for statement speech length. Utterance content was also explored.

Results: Clustering by patient explained 30% of the variance in utterances *f0*. Cues and concerns were on average 13.07 Hz higher than neutral statements (p = 0.02). Cues and concerns in these consultations contained content with a high proportion of recurrence fears.

Conclusion: The present study highlights the benefits and challenges of adding *f0* and potential other prosodic features to the toolkit of coding emotional distress in the health communication setting.

Practice implications: The assessment of *f0* during clinical conversations can provide additional information for research into emotional expression.

Abbreviations

HNC	Head and Neck Cancer
FoR	Fears of cancer Recurrence
VR-CoDES	Verona Coding Definitions of Emotional Sequence
BSP	Behavioural Signal Processing
ASP	Affective Signal Processing
F0/f0	Fundamental Frequency of Pitch (Hz)
QoL	Quality of Life

Keywords:. Emotional distress, The VR-CoDES, Fundamental frequency of pitch , Multilevel , Head and Neck cancer

1. **Introduction**

Previous studies have shown that health provider responses to expressions of emotion can depend on, for example, who initiated the disclosure, with patients' initiations more often being shut down [1, 2]. When and the way in which concerns are presented during the consultation can also impact whether they are met by empathy and given space, or shut down [2-4]. Furthermore, Kennifer *et al*. [5] found that the intensity and type

of emotional concern (e.g. anxiety or sadness) impacts how, and if at all, the medical consultants respond.

Due to the significant impact this can have on patients' quality of life (QoL), guidelines for patient-centered care call for these concerns to be responded to with empathy [6-8] and signposting to further psychological and other support services where appropriate [9]. Improvements have been accomplished by developing additional guidelines and communication training in how to detect and manage emotional distress for care providers [4, 10]. Advances in the area are aided by the development of research coding tools, used to identifying and classify patient concerns and consultant responses [11].

Our established research team, decided to explore the potential for intensive investigation of emotional expression in patients attending a tertiary surgical/oncology centre. We were motivated to conduct our research with clinical interactions, in the oncology field, that are frequently interspersed with emotional content, as has been previously shown by our group [3] and others [2]. Patients, in cancer clinics especially, may hide their emotions through embarrassment, a wish not to burden others, accepting anxiety as part of treatment or because they were not asked about additional concerns [12]. Hence emotional arousal within the cancer consultation may not always be easy to identify and therefore it can be argued that this setting warrants focussed attention by researchers.

1.1. Fundamental frequency of pitch (*f0*)

Reviewing the literature on vocal characteristics of human affect, fundamental frequency of pitch (*f0*), has been identified as one of the most reliable tools, essential for detecting emotional arousal using the voice [13-19]. However, no study has yet used this objective measure of emotional distress to characterise patient concern in the clinical oncology setting.

Helmholtz' concept, that characteristics of the voice reflect an individual's state of mind, may be the starting point of research into what information is carried by speech beyond semantic meaning (Helmholtz as cited in [20]). One of the current goals of behavioural signal processing (BSP) is to combine human and machine observations of dyadic interactions to develop computational models that can assess social, emotional, and communicative states of interlocutors in great detail [21]. Beginning to integrate and test the applicability of these approaches to the study of health communication could therefore lead to great benefits, particularly when studying the vocal properties of empathy [22, 23]

Kirshnan and Fernandez's methods for recognising emotional states from speech indicate the importance of *f0* (U.S. Patent No. 8,595,005 B2, 2013) [24]. Affective speech processing (ASP) studies are advancing automatic methods to identify emotions in speech using *f0* [25]. Higher levels of emotional arousal (i.e. anxiety, distress) can be identified as they are associated with higher *f0* within the speaker's vocal range and context [16, 26]. Furthermore, *f0* is an advantage over other behavioural signals, as

it is a nonintrusive tool and extracting vocal arousal can assess therapist's empathy through their relational pitch [22] [23]. A systematic search of 'Web of Science' and 'PubMed' databases demonstrated that *f0* has never been used in oncology before.

1.2. The Verona Coding Definitions of Emotional Sequences

The extensive work of Scherer has developed a highly generalizable model of conceptualising emotions in vocal communication. He has adapted the Brunswik lens model approach of human perception to vocal expression and proposed an important extension known as the tripartite emotion expression and perception (TEEP) model [27]. It states that it is valuable to conceive the emotion expression from the person, in this case, the patient, from a distal acoustic analysis (e.g. *f0*) and encoded via proximal vocal cues, and determine these cues using subjective ratings from naive observers. The proximal or decoding part of the vocalisations can be elegantly identified using the Verona Coding Definitions of Emotional Sequences (VR-CoDES), which were developed specifically to discern occurrences of emotional distress from both verbal and non-verbal content in medical consultations [11]. This coding scheme guides researchers to identify implicit and explicit expressions of negative emotions in patients, labelled as different cues and concerns respectively [11]. Furthermore, they provide a uniform coding system for health provider responses [28] . The VR-CoDES is a widely

accepted method, whose validity and inter-coder reliability has been tested on many occasions (e.g. [3]).

1.3. Combining the two approaches

This study will pioneer the application of intonation analysis to utterances identified by VR-CoDES and test the ability to apply this method in an ecological setting, i.e. a busy, highly medically focused outpatient session. When using the VR-CoDES coding scheme to evaluate the quality of varying communication in the health care system, it is essential that instances can be classified depending on the intensity of emotional arousal as they may, or may not, warrant different responses. This study, by combining the VR-CoDES with *f0* analysis, will enable the validation of the coding systems' goal of detecting emotional distress using an objective physiological measurement. It is accepted that the VR-CoDES does include specific reference to non-verbal cues including 'tone of voice' however the inclusion of assessing *f0* can support the coding system's value placed on the 'tone of voice' interpretation by potentially acting as an adjunct for the presence of prosodic elements within the VR-CoDES definitions.

1.4. Aims

This study has two aims. First, to investigate the feasibility of using *f0* to classify the presence of emotional distress of cancer patients during follow-up consultations. Second, to test for an association of emotional

arousal (*f0)* and utterances selected by the VR-CoDES guidelines. It is hypothesised that identified cues and concerns using the VR-CoDES system will have significantly higher levels of *f0* than randomly selected neutral patient speech.

2. Methods

2.1. Participants and Procedures

Any patients who had a previous diagnosis of head and neck (HNC) and had completed primary treatment for at least 6 weeks, were disease free and under a routine surveillance programme at Aintree University Hospital Maxillofacial Surgery (Liverpool, UK), were included in recruitment by the surgery team. Recruitment was held over four month's duration. Previous work by our team has shown that emotions are detectable in clinical interactions [3]. Our intention was to collect a minimum of 30 patients which we estimated would provide approximately 100 cues and concerns (calculated from previous study)[3]. No formal power analysis was attempted as neither the variance nor Intra-Class Correlation of *f0* across patients were known. Patients in palliative care or attending the clinic for other medical reasons were excluded. Forty-one patients were approached. Nine patients refused (83% response rate). Subsequently, the follow-up outpatient consultations of 32 HNC survivors were audio-recorded using a Tascam DR-40 microphone (WAV: 44.1kHz/48kHz/96kHz, 16 bit, Stereo).

After controlling for quality of sound files and excluding any consultations with excessive background noise (n= 3), consultations without emotional distress cues (n = 6), as well as patients with speaking difficulties / low speech quality (n = 7); 16 patient consultations were included in the analysis. All recordings were made with the same consultant (Professor Dr Rogers) to limit variability. Data were stored and processed securely, protecting patient confidentiality by removing references to patient identifiers. All participants gave informed written consent..

The study used an exploratory, non-randomised, mixed design to investigate the pitch characteristics of patients' expressions of emotional distress. It gained the ethical approval on the 30th March 2015, by the North West 3 Research Ethics Committee – Liverpool East (R&D ref: 797/15; REC Ref: 15/NW/0173).

2.2. Outcome Measures

Patient cues and concerns [11] and consultant responses [28] were coded using the VR-CoDES via the Observer® XT 12.0 system for Windows. See Zhou *et al*. [3] for examples of cues and concerns. The number, type and time stamp of patient emotional cues and concerns were noted during coding to enable the subsequent identification of pitch of these utterances. Coding was completed by two postgraduate students trained in VR-CoDES and inter- and intra-rater reliability [29] were satisfactory (see Table 1).

The prosodic parameter 'intonation' is the change in pitch (*f0)* over time and is only meaningful for interpretation when placed in the context

(pitch range and register) of the speaker [30]. The auditory perception of high or low pitch is caused by the frequency with which the speaker's vocal folds in the larynx open and close: the faster they move, the higher the pitch [30]. *F0* measures the complete number of cycles within one second of speech and is measured in Hertz (Hz) [20]. An individual's speech using 200 complete cycles per second is hence said to have a pitch of 200 Hz. *F0* was therefore interpreted in light of the speaker's individual range, by collecting neutral utterances for comparison.

The type of emotion present (i.e. anxiety or sadness) also impacts *f0* (U.S. Patent No. 8,595,005 B2, 2013) and *f0* discriminates between high and low arousal emotions but is not necessarily sensitive to whether the arousal is a positive or negative emotion [26]. The definition of 'neutral' statements for this study subsequently only included any non-emotional patient speech (excluding laughter etc.) lasting at least 1 second. Furthermore, any cues, concerns, and neutral statements where background chatting or noise occurred had to be excluded.

Audio editor Audacity (http://audacityteam.org/) was used to slice recordings into units of analysis as outlined by the time stamps set out by VR-CoDES definitions [11], as well as to cut out any instances of overlapping speech or background noise still remaining. Segment lengths ranged from 0.75 to 18.25 seconds. The weight of the utterance (speech length to the closest quarter second) was taken into consideration for the analysis. The number of cues/concerns and neutral statements for each patient were

balanced as best as possible. If an abundance of neutral statements were present compared to cues and concerns for one patient, they were chosen at random. Following this, the prosody of the unit chunks was analysed using PRAAT 5.3.51 software [31] using a script developed by Imel *et al*. [22] which calculated the mean *f0* every 0.25 seconds and a band pass filter of 75 to 500 Hz was set to accommodate the range of human speech.

2.3. Data Analysis

A hierarchical linear model with random intercept and two-level covariates was chosen for statistical analysis as it is robust against unbalanced data, missing data points and differing utterance lengths [32]. Observations were not independent of each other, which was controlled for by nesting utterances (level 1) within patients (level 2). The outcome variable was mean *f0*. First a random effects ANOVA to determine the variance at the patient clustering level was conducted. Secondly, further explanatory variables were entered into the model. At level 1, this consisted of the type of utterance (cue/concern 1; neutral 0) and clip length (duration of speech to the nearest 0.25 second) and at level 2 covariates included patient age and gender (male 1, female 0). The analysis was conducted using STATA/IC™ 13.0 for Windows using the 'xtmixed' procedure (e.g. [33]).

Graphs of the utterances were also created to demonstrate the pitch trajectories measured and their scales gender adjusted (100-300 Hz for women, 50-150 Hz for men) [23] to facilitate comparisons and qualitative analysis, including content exploration (for examples see Figures 1 and 2).

3. Results

3.1. Participants

Of the 16 HNC survivors included in the study, the mean (SD) age was 62.75 (8.07) years (range 52-79) and 62.5% (n = 10) of participants were male. Diagnoses ranged from seven (43.8%) patients with stage 1, one patient (6.25%) with stage 2, two (12.5%) with stage 4a**,** two (12.5%) with neck nodes, two (12.5%) with sublingual/parotid glands and two (12.5%) unstaged tonsillar tumours. Primary treatment consisted of surgery alone (n = 6, 37.5%), surgery and adjuvant radiotherapy (n=9, 56.3%) or chemo-radiotherapy alone (n=1, 6.3%) and average time since diagnosis was 2 years and 8 months.

3.2. Emotional arousal and the VR-CoDES

Each utterance clip (n=89) was clustered by patient (n=16). The clustering size varied from a minimum of 2 to a maximum of 12, with an average of 5.6 clips per patient. The clips consisted of 44 cues and concerns, and 45 neutral statements, with the VR-CoDES consisting of 5 concerns, 25 type 'b' cues (verbal hints to hidden concern) and 14 other cues.

A log transformation of means was conducted to adjust for skewness and heteroscedasticity. Both the log transformed and the raw data resulted in significant differences in mean *f0* associated with coded cues and concerns versus neutral utterances using a hierarchical linear model. Log transformation displayed a slight improvement in the fit of the model (LR

test vs. linear regression: 20.84 versus 13.95, both at p<0.001), however, for ease of interpretation raw data were used to display findings.

The grand mean across all utterances was 121.32 Hz. After conducting the random effects ANOVA, the estimate of the variance component of patient clustering was 310.12 Hz and the likelihood ratio (LR) test statistic for the null hypothesis, that there is no cross-patient variation in pitch, was 24.50 with one degree of freedom (p<0.001), providing evidence that there is a very high degree of inter-dependence of the data, making clustering by patient essential. Intra-class correlation was calculated to be 0.304 and therefore about 30% of variation can be explained by differences across patients, with 70% of variation left to be explained by within patient differences [33].

Adding level-1 and level-2 covariates to the model, type of utterance (cue/neutral) and gender had a significant effect, while clip length and patient age did not (Table 2).

The Wald test found that the subset of coefficients were jointly significant (W = 9.58, p = 0.048). Variation in intercepts was substantial at 200.72 and, as mentioned above, the fit of the random intercept model was preferable to that of the regression model (LR = 13.95, p < 0.001). Together the significant gender and cue/neutral covariates, therefore, accounted for 15% of the variation in intercept. Utterances from males, compared to females, were expected to have a pitch of 19.34 Hz lower (p = 0.04). Compared to neutral statements, the pitch of VR-CoDES cues or concerns

was on average 13.07 Hz higher (p = 0.023) (Table 2), see Figures 1 and 2 for examples. These findings, established using data from an ecological setting, support the hypothesis that there is an association between heightened vocal arousal (*f0*) and VR-CoDES cues and concerns.

3.3. Qualitative Analysis

The content of each utterance was explored by splitting them into consultations which displayed the pattern of having cues and concerns with higher average *f0* as compared to their neutral statements and consultations where the *f0* of cues and concerns was equal or lower than its neutral statements (see Table 3). As Table 3 portrays, VR-CoDES which displayed the positive association consisted mainly of recurrence worries or severe pain and physical problems. In contrast, the content of VR-CoDES that did not follow the association consisted mainly of difficulties with eating, with only a couple of instances of apprehension about reoccurrence, and overall less intense concerns, adding support to the use of *f0* when identifying heightened emotional arousal.

4. Discussion and conclusion

4.1. Discussion

This study extracted *f0* as an objective pitch feature to characterise the distress in VR-CoDES cues and concerns using the audio recordings of

HNC survivors' follow-up consultations. In comparison to the values of neutral statements, there was a significant difference in the pitch of statements classified as cues and concerns by the VR-CoDES. The statistical difference, higher *f0* averages, indicates a trend consistent with the literature that individuals use an increase in the pitch of their voice when speaking with anxiety or other emotional arousal [13-19]. It is difficult to assign a level that would be considered a clinical difference. However, it is interesting that the qualitative analysis indicated that heightened pitch was associated more closely with instances related to FoR than it was with problems regarding eating, which supports the literature that FoR is a major source of distress for HNC survivors [34, 35].

The goal of the VR-CoDES coding scheme is to facilitate comparative research of communication between a health care provider and patient, during which the latter expresses emotional distress. Using definitions based upon qualities of speech, it guides users to pick out moments of concern, uncertainty, and unrest that call for a supportive response from the health care provider, even when these are not directly expressed [11]. The results of this study indicated that *f0* may be an additional objective tool with which instances of distress can be classified. Simultaneously, the findings support the VR-CoDES and the system's objective to include instances of emotional expression identified not only by using verbatim data, but also by listening to the way in which words are said, to assist with identifying emotions. The rise in emotional arousal, measured using *f0,* was significant despite the

majority of instances included in this study consisting of 'hidden' emotions (i.e. cues).

A main strength of this study lies in its ecological setting, however, there are several considerations to be made when using *f0* to classify the presence of emotional distress in busy HNC follow-up consultations. The importance of nesting each pitch value within the individual patient was demonstrated by the high degree of inter-dependence of the data. This confirms the literature that each individual's unique range is significantly different, with each gender's range varying distinctively [30] [23]

In experimental settings, the type of emotion being expressed can often be closely defined and controlled [20]. As previously mentioned, the definitions set out by VR-CoDES (an 'unpleasant emotion' that may warrant clarification) [11] (p147) seems to closely match the description of 'emotional arousal', measured by *f0*. Although the results display a significant association, this wasn't true in all cases and could potentially be explained by discrepancies in what each measure target**s**. The VR-CoDES does not differentiate between different types of emotion and as the content exploration revealed, this offers an explanation for why findings were not completely consistent in this particular data set. Therefore, the need for clear and detailed definitions of the type of emotion the investigator is seeking to detect is recommended to derive universal conclusions. Prosodic features may be able to add to the development of a novel classification system, multi-dimensional in its nature, as has been

suggested [14]. Scherer has identified additional features such as 'uptrend *f0* contour', that is rising intonation, may improve identification of emotional speech. Hence, the configuration of speech intonation may also be a key feature [36]. In summary, this study's findings tentatively indicate that the height of *f0* may help explore heightened emotional arousal that is related to specific topics, such as FoR in this study.

It is recognised that *f0* alone is probably not sufficient. However, this study has been valuable in highlighting the possibility of combining lexical VR-CoDES with pitch features to develop a multilevel approach. The digital resynthesis of speech prosody has shown, previously, that the influence of speaker and utterance content was of minor influence on naive judges' assessments of emotions, which indicates that these results may generalise somewhat [37].

The length for one unit of analysis cannot be chosen in an ecological setting, yet, longer segments of uninterrupted speech may have an increased chance of containing multiple emotional characteristics [38]. Previous studies have adapted their data to adjust for this problem e.g. [39]. This study controlled for speech length by adding it into the multilevel model as a covariate and found that it had no significant impact on *f0*.

Overall, the results confirm that it is possible to determine the presence of emotional arousal in VR-CoDES using audio-recordings from HNC consultations. Similarly, anxiety (high arousal) has been found to be higher during pre-treatment versus follow-up clinics, as used for this study

[40]. Pitch therefore seems to be a sensitive enough measure for circumstances with comparatively lower levels of anxiety and it may be beneficial to add this measure to other coding schemes.

Inevitably, there are some limitations to the approach presented. First, the immediate context, preceding the speech whose *f0* is measured, is very important since an individual's voice interacts with its surroundings [30]. For example, during their study of therapist empathy, Xiao *et al*. [41] established a negative correlation with the prosodic features: energy and pitch. This relationship was still present, but reduced in its significance when taken out of the context of the previous patient's speech turn. The association between VR-CoDES and heightened *f0* was determined by nesting them within each speaker's voice range using neutral statements, however, not within the context of any preceding behaviour. This was due to the methods available and inevitable difficulties with overlapping speech and background noise, as well as variations in who was speaking in the preceding turn. Analysing the distribution of preceding vocal patterns and nesting them within their own context may be beneficial in showing clearer associations.

A further limitation to the study was the inability to include every patient, such as individuals who had treatment affecting their tongue or that limited the clarity of their speech, e.g. tumours surrounding the larynx [42, 43]. In addition, more severely ill individuals may have been more likely to refuse consent and recording as they may have difficulties expressing

themselves. However, there was no indication that the type of treatment received by patients impacted on anxiety or FoR levels during follow-up sessions.

Finally, researchers in healthcare communication may not be familiar with handling the comparatively large set of data points that are derived from extracting *f*0 values from speech. We believe that this is an issue of familiarity and that with experience the researcher will find the extra information gleaned from this approach to be worthwhile.

Integrating *f0* into the classification of distress may in future be useful for differentiating between concerns that are highly charged in emotional arousal and those that are focused on garnering information. Depending on this element, different responses may be warranted to lead to the best patient care. As previous studies have found, much value also lies in looking at prosodic features of consultant responses as they can reveal the degree of empathy expressed [22, 23, 39, 41] which in turn affects patients' satisfaction (e.g. CARE measure rating, [44]). Future work in prosodic recognition may therefore lead to detailed feedback regarding the 'quality' of consultant responses.

4.2. Conclusion

This study was the first attempt to code emotional arousal in the oncology setting. Although not yet robust and presented with many challenges, there is a significant association between VR-CoDES and higher levels of pitch. Indications of higher emotional arousal attached to concerns

regarding FoR were also found. Future research into the integration of prosodic measures into health communication research in oncology are warranted, as objective characteristics of emotional intensity may be of value for many goals in this field.

4.3. Practice implications

The assessment of f0 during clinical conversations can provide additional information that can be utilised for research into emotional expression. This methodology has been applied using significant sample sizes to make comparisons between two cognitive therapy interventions for couple therapy. [22] Hence modern computer technology can now manage sizeable data sets with comparative ease. The next steps in the application of *f*0 is to use the assessment as an adjunct to VR-CoDES in further studies. Continuously improving the robustness of the measure, including its application in other settings, could lead to the development of a universal rating scale screening tool of the level of emotional arousal of patients overall, as well as attached to the issues they discuss. It could also add to strategies for how consultants can improve recognition, exploration, and therapeutic action towards emotional stress, such as those discussed by the 3-stage model of patient-centred communication for addressing cancer patients' emotional distress [45].

Conflicts of interest

There were no conflicts of interest.

Acknowledgements

Support was received from Aintree University Hospital NHS Foundation Trust for recruitment, data collection, and the Ethics application process. Generous assistance in script writing for PRAAT was received by Brian Baucom from University of Utah.

References

[1] Del Piccolo L, Mazzi MA, Goss C, Rimondini M, Zimmermann C. How emotions emerge and are dealt with in first diagnostic consultations in psychiatry. Patient Educ Couns. 2012;88:29-35.

[2] Finset A, Heyn L, Ruland C. Patterns in clinicians' responses to patient emotion in cancer care. Patient Educ Couns. 2013;93:80-5.

[3] Zhou Y, Humphris G, Ghazali N, Friderichs S, Grosset D, Rogers SN. How head and neck consultants manage patients' emotional distress during cancer follow-up consultations: a multilevel study. Eur Arch Otorhinolaryngol. 2015;272:2473-81.

[4] Zimmermann C, Del Piccolo L, Finset A. Cues and concerns by patients in medical consultations: a literature review. Psychological bulletin. 2007;133:438-63.

[5] Kennifer SL, Alexander SC, Pollak KI, Jeffreys AS, Olsen MK, Rodriguez KL, et al. Negative emotions in cancer care: do oncologists' responses depend on severity and type of emotion? Patient Educ Couns. 2009;76:51-6.

[6] Perocchia RS, Hodorowski JK, Williams LA, Kornfeld J, Davis NL, Monroe M, et al. Patient-centered communication in cancer care: the role of the NCI's Cancer Information Service. J Cancer Educ. 2011;26:36-43.

[7] Stewart M, Brown JB, Donner A, McWhinney IR, Oates J, Weston WW, et al. The impact of patient-centered care on outcomes. The Journal of family practice. 2000;49:796-804.

[8] Epstein R, Street R. Patient-centered communication in cancer care: Promoting healing and reducing suffering. Bethesda, MD: National Cancer Institute, National Institutes of Health; 2007.

[9] Ghazali N, Kanatas A, Langley DJ, Scott B, Lowe D, Rogers SN. Treatment referral before and after the introduction of the Liverpool Patients Concerns Inventory (PCI) into routine head and neck oncology outpatient clinics. Support Care Cancer. 2011;19:1879-86.

[10] Kanatas A, Ghazali N, Lowe D, Rogers SN. The identification of mood and anxiety concerns using the patients concerns inventory following head and neck cancer. Int J Oral Maxillofac Surg. 2012;41:429-36.

[11] Zimmermann C, Del Piccolo L, Bensing J, Bergvik S, De Haes H, Eide H, et al. Coding patient emotional cues and concerns in medical consultations: the Verona coding definitions of emotional sequences (VR-CoDES). Patient Educ Couns. 2011;82:141-8.

[12] Dean C, Surtees P. Do psychological factors predict survival in breast cancer? Journal of Psychosomatic Research. 1989;33:561-9.

[13] Juslin P, Scherer K. Vocal expression of affect. In: Harrigan J, Rosenthal R, Scherer K, editors. The New Handbook of Methods in Nonverbal Behavior Research. Oxford, UK: Oxford University Press; 2005.

[14] Grimm M, Kroschel K, Mower E, Narayanan S. Primitives-based evaluation and estimation of emotions in speech. Speech Commun. 2007;49:787-800.

[15] Bulut M, Narayanan S. On the robustness of overall F0-only modifications to the perception of emotions in speech. The Journal of the Acoustical Society of America. 2008;123:4547-58.

[16] Busso C, Bulut M, Lee S, Narayanan S. Fundamental frequency analysis for speech emotion processing. In: Hancil S, Lang P, editors. The Role of Prosody in Affective Speech. Berlin, Germany: Publishing Group; 2009.

[17] Lee C-C, Mower E, Busso C, Lee S, Narayanan S. Emotion recognition using a hierarchical binary decision tree approach. Speech Communication. 2011;53:1162-71.

[18] Yildirim S. Detecting emotional state of a child in a conversational computer game. Computer speech & language. 2011;25:29-44.

[19] Black MP, Katsamanis A, Baucom BR, Lee C-C, Lammert AC, Christensen A, et al. Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. Speech Commun. 2013;55:1-21.

[20] van den Broek EL. Emotional Prosody Measurement (EPM): a voice-based evaluation method for psychological therapy effectiveness. Studies in health technology and informatics. 2004;103:118-25.

[21] Narayanan S, Georgiou PG. Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language: Computational techniques are presented to analyze and model expressed and perceived human behavior–variedly characterized as typical, atypical, distressed, and disordered–from speech and language cues and their applications in health, commerce, education, and beyond. Proceedings of the IEEE Institute of Electrical and Electronics Engineers. 2013;101:1203-33.

[22] Imel ZE, Barco JS, Brown HJ, Baucom BR, Baer JS, Kircher JC, et al. The association of therapist empathy and synchrony in vocally encoded arousal. Journal of counseling psychology. 2014;61:146-53.

[23] Weiste E, Perakyla A. Prosody and empathic communication in psychotherapy interaction. Psychotherapy research : journal of the Society for Psychotherapy Research. 2014;24:687-701.

[24] Kirshnan A, Fernandez M. In: Office PaT, editor. U.S.2013.

[25] Arias J, Busso C, Yoma N. Shape-based modeling of the fundamental frequency contour for emotion detection in speech. Computer Speech and Language. 2014;28:278-94.

[26] Guidi A, Vanello N, Bertschy G, Gentili C, Landini L, Scilingo E. Automatic analysis of speech F0 contour for the characterization of mood changes in bipolar patients. Biomedical Signal Processing and Control. 2015;17 29-37.

[27] Banziger T, Hosoya G, Scherer KR. Path Models of Vocal Emotion Communication. PloS one. 2015;10:e0136675.

[28] Del Piccolo L, de Haes H, Heaven C, Jansen J, Verheul W, Bensing J, et al. Development of the Verona coding definitions of emotional sequences to code health providers' responses (VR-CoDES-P) to patient cues and concerns. Patient Educ Couns. 2011;82:149-55.

[29] Altman D. Practical statistics for medical research. London: Chapman and Hall; 1991.

[30] Szczepek Reed B. Analysing Conversation: An Introduction to Prosody. New York, NY: Palgrave Macmillan. ; 2011.

[31] Boersma P, Weenink D. Praat: doing phonetics by computer. 5.3.51 ed. http://www.praat.org/2013.

[32] Rabe-Hesketh S, Skrondal A. Multilevel and Longitudinal Modeling Using Stata. 3rd ed. College Station, Texas: Stata Press; 2012.

[33] Steenbergen M. Hierarchical Linear Models for Electoral Research: A Worked Example in Stata. 2012.

[34] Humphris GM, Rogers S, McNally D, Lee-Jones C, Brown J, Vaughan D. Fear of recurrence and possible cases of anxiety and depression in orofacial cancer patients. Int J Oral Maxillofac Surg. 2003;32:486-91.

[35] Rogers SN, El-Sheikha J, Lowe D. The development of a Patients Concerns Inventory (PCI) to help reveal patients concerns in the head and neck clinic. Oral Oncol. 2009;45:555-61.

[36] Scherer K, Johnstone T, Klasmeyer G. Vocal Expression of Emotion. In: Davidson R, Scherer K, Goldsmith H, editors. Handbook of Affective Sciences. USA: Oxford University Press; 2002. p. 433-56.

[37] Ladd D, Silverman K, Tolkmitt F, Bergmann G, Scherer K. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. Journal of the Acoustical Society of America. 1985;78:435-44.

[38] Schuller B, Batliner A, Steidl S, Seppi D. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. Speech Commun. 2011;53:1062-87.

[39] Lee C-C, Black M, Katsamanis A, Lammert A, Baucom B, Christensen A, et al. Quantification of Prosodic Entrainment in Affective Spontaneous Spoken Interactions of Married Couples2010.

[40] Kanatas A, Humphris G, Lowe D, Rogers SN. Further analysis of the emotional consequences of head and neck cancer as reflected by the Patients' Concerns Inventory. Br J Oral Maxillofac Surg. 2015;53:711-8.

[41] Xiao B, Bone D, Van Segbroeck M, Imel Z, Atkins D, Georgiou P, et al. Modeling Therapist Empathy through Prosody in Drug Addiction

Counselling. 15th Annual Conference of the International Speech Communication Association. Singapore2014.

[42] Hoyt DJ, Lettinga JW, Leopold KA, Fisher SR. The effect of head and neck radiation therapy on voice quality. Laryngoscope. 1992;102:477-80.

[43] Verdonck-De Leeuw I, Koopmans-Van Beinum F. The effect of radiotherapy on various acoustical, clinical and perceptual pitch measures. XIIIth International Congress of Phonetic Sciences. Stockholm1995. p. 610-3.

[44] Mercer SW, Maxwell M, Heaney D, Watt GC. The consultation and relational empathy (CARE) measure: development and preliminary validation and reliability of an empathy-based consultation process measure. Family practice. 2004;21:699-705.

[45] Dean M, Street RL, Jr. A 3-stage model of patient-centered communication for addressing cancer patients' emotional distress. Patient Educ Couns. 2014;94:143-8.

List of Legends for Figures and Tables

Figure 1a –

PRAAT *f0* pitch graph (Hz) of a neutral statement: average pitch 118.22 Hz, speech length 4.65 sec (participant 36)

Figure 1b –

PRAAT *f0* pitch graph (Hz) of a cue statement: average pitch 134.27 Hz, speech length 4.5 sec (participant 36)

Figure1a

Figure 1a (this text box to be erased, it is simply a label)

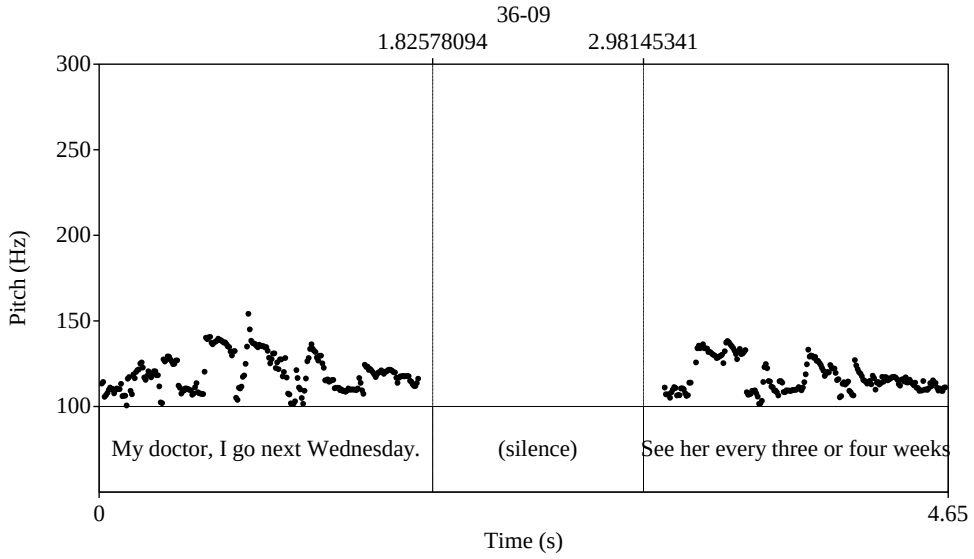
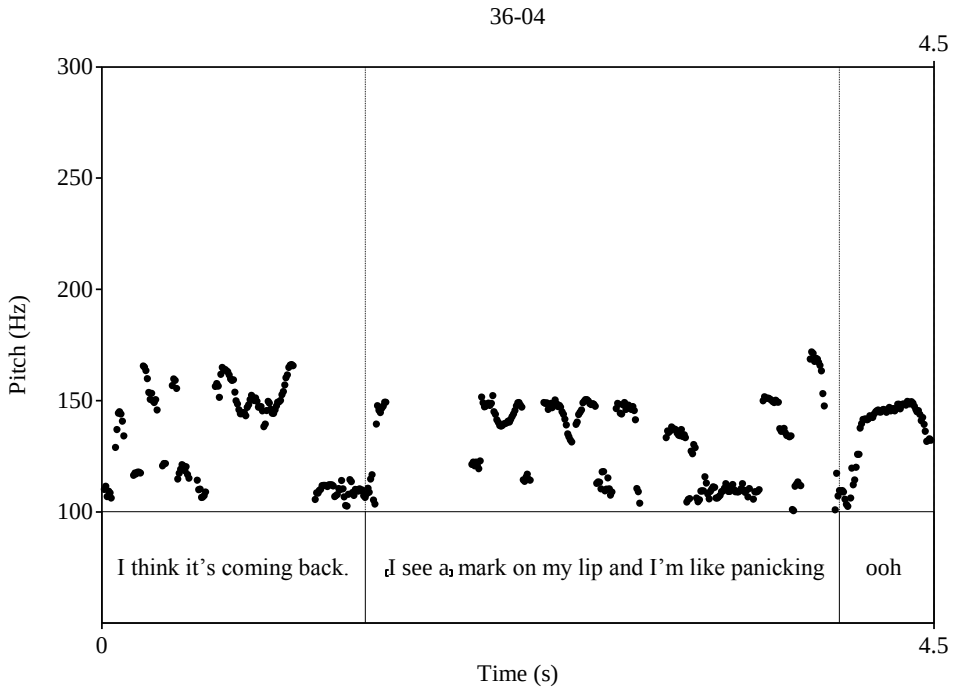


Figure1b

Figure 1b (this text box to be erased, it is simply a label)



(For Tables see Excel sheets)

Table 1 - Results of inter- and intra-rater reliability for frequency and sequence of cue/concerns and responses (1 sec tolerance window).

Type	**Occasion of check**	**No. of transcripts**	**Cohen's Kappa (95% CI)**	**Agreement (%)**
Inter-rater (Coders 1&2)	Close to the end	2	0.79 (0.63, 0.94)	82.25
Intra-rater	Coder 1	1	0.78 (0.63, 0.94)	82.86
	Coder 2	1	0.77 (0.58, 0.97)	80

Table 2 - Hierarchical linear model, using raw data estimates, for the outcome variable mean f0.

Covariate	**Coefficient**	**Standard Error**	**Z**	**P>|z|**	**[95% Conf. Interval]**
Clip length	0.52	1.08	0.48	0.63	-1.59, 2.64
Cue 1/Neutral 0	13.07	5.74	2.28	0.023	1.82, 24.31
Age	0.35	0.61	0.58	0.57	-0.84, 1.53
Sex	-19.34	9.48	-2.04	0.041	-37.93, -0.76
Constant	103.41	39.13	2.64	0.008	26.70, 180.11
Random-effects Parameters		**Estimate**	**Standard Error**	**[95% Conf. Interval]**	
Patient Number: Identity					
SD (Constant)		14.2	3.63	8.61, 23.42	
SD (Residual)		25.96	2.09	22.18, 30.39	

Table 3 – Content of VR-CoDES cues and concerns, by patients and on whether VR-CoDES in the consultation displayed the trend of heighted f0 values compared to neutral statements.

Patient ID	**VR-CoDES content of consultations whose average VR-CoDES *f0* was higher than neutral statements**	**Patient ID**	**VR-CoDES content of consultations whose average VR-CoDES *f0* was the same/lower than neutral statements**
11 (M)	Cancer recurrence (cue b x2)	10 (M)	Eating problems (cue c, b)
12 (F)	Severe pain, inability to walk, financial difficulties (cue b x6, e x2, d x2)	14 (M)	Eating, swallowing problems (cue c, d)
15 (M)	Anxiety about cancer recurrence, slight hearing problem (concern x2, cue b x2)	16 (M)	Eating, swallowing problems, some worry of recurrence (cue b x2, e, concern)
18 (M)	Scarring (cue b)	17 (M)	Infection (concern)
25 (F)	Eating problem, tightness, Pain (cue b x2)	24 (F)	Eating problems, taste (cue b x2, e)
29 (F)	Down about eating problems (cue b)	28 (M)	Apprehension about appointment / recurrence (cue b, a x2, g x2)
36 (F)	Lip discomfort, panic about recurrence (cue b x2)	33 (M)	Apprehension about possible future pain (concern)
38 (M)	Apprehension about recurrence (cue d, b)	35 (F)	Pain on mouth (cue b, concern)