



**HAL**  
open science

# Quietly Angry, Loudly Happy: Self-Reported Customer Satisfaction Vs. Automatically Detected Emotion In Contact Center Calls

Eric Bolo, Muhammad Samoul, Nicolas Seichepine, Mohamed Chetouani

► **To cite this version:**

Eric Bolo, Muhammad Samoul, Nicolas Seichepine, Mohamed Chetouani. Quietly Angry, Loudly Happy: Self-Reported Customer Satisfaction Vs. Automatically Detected Emotion In Contact Center Calls. *Interaction Studies*, 2023, 24 (1), pp.168-192. 10.1075/is.22038.bol . hal-04109350

**HAL Id: hal-04109350**

**<https://hal.science/hal-04109350>**

Submitted on 30 May 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# QUIETLY ANGRY, LOUDLY HAPPY: SELF-REPORTED CUSTOMER SATISFACTION VS. AUTOMATICALLY DETECTED EMOTION IN CONTACT CENTER CALLS

---

VIHAR 2021 SPECIAL ISSUE, INTERACTION STUDIES

**Eric Bolo**  
Batvoice AI, Paris, France  
ebolo@batvoice.com

**Muhammad Samoul**  
Batvoice AI, Paris, France  
msamoul@batvoice.com

**Nicolas Seichepine**  
Batvoice AI, Paris, France  
nseichepine@batvoice.com

**Mohamed Chetouani**  
ISIR, CNRS UMR 7222,  
Sorbonne University, France  
and Batvoice AI, Paris, France  
mohamed.chetouani@sorbonne-universite.fr

## ABSTRACT

1 Phone calls are an essential communication channel in today's contact centers, but they are more  
2 difficult to analyze than written or form-based interactions. To that end, companies have tradi-  
3 tionally used surveys to gather feedback and gauge customer satisfaction. In this work, we study  
4 the relationship between self-reported customer satisfaction (CSAT) and automatic utterance-level  
5 indicators of emotion produced by affect recognition models, using a real dataset of contact center  
6 calls. We find (1) that positive valence is associated with higher CSAT scores, while the presence  
7 of anger is associated with lower CSAT scores; (2) that automatically detected affective events and  
8 CSAT response rate are linked, with calls containing anger/positive valence exhibiting respectively a  
9 lower/higher response rate; (3) that the dynamics of detected emotions are linked with both CSAT  
10 scores and response rate, and that emotions detected at the end of the call have a greater weight in the  
11 relationship. These findings highlight a selection bias in self-reported CSAT leading respectively to  
12 an over/under-representation of positive/negative affect.

13 *Keywords* Customer satisfaction · Emotions · Affective Computing · Real-World applications

## 14 1 Introduction

15 In spite of digitalization, phone calls remain a major communication channel in today's contact centers, but they are more  
16 difficult to analyze than written or form-based interactions. In an increasingly customer-centric business environment,  
17 gathering insight from interactions has become a common practice. To that end, companies have traditionally used  
18 surveys to gather feedback and evaluate service quality using metrics such as Customer Satisfaction (CSAT). A CSAT  
19 form asks the customer to assess their level of satisfaction on a Likert-scale. In the context of call centers, the question  
20 is typically asked by the operator at the end of the call.

21 Even if widely used, CSAT presents important limitations: first, the response rate is usually well below 100%, pointing  
22 to a probable selection bias, as the population of responders is likely to differ from the overall population; second, it is a  
23 summary metric which cannot capture the full complexity of customer experience, nor reveal details about its evolution.

24 In light of these limitations, it is natural to look for alternative or complementary metrics which (1) are objective rather  
25 than declarative, (2) can be applied on all interactions rather than a biased sample, and (3) give a higher-resolution  
26 description of each customer's experience. AI-based approaches are promising with respect to all three criteria, as

27 they do not depend on explicit feedback, can be applied on all interactions or a random sample, and can yield useful  
28 information down the level of an utterance and sometimes further.

29 In this work, we study the relevance of affective computing techniques for evaluating service quality and customer  
30 satisfaction. More precisely, we study the relationship between automatic utterance-level indicators of emotion and  
31 CSAT, using a real-world dataset of call center interactions.

32 For this study, data was anonymized by detecting identifying information in transcripts produced by an Automatic  
33 Speech Recognition (ASR) system and removing corresponding audio segments. Only customer speech turns were  
34 taken into account, and all identifying metadata concerning customers and operators were discarded.

35 More specifically, we investigate the relationship between affective indicators and satisfaction by analyzing CSAT  
36 response rate, proportions of high satisfaction and proportions of affective events in calls. For this purpose, we conducted  
37 several analyses using affective computing models designed to predict positive/negative valence and detect anger, which  
38 are used to label a very large dataset of real calls (160 630 calls). We find that positive valence and anger as well as  
39 their dynamics in each call are meaningfully linked with both CSAT score and response rate.

## 40 1.1 Contributions

41 The main contributions of this work are as follows:

- 42 • we jointly study automatically detected emotions and self-reported customer satisfaction using a large dataset  
43 of contact center calls (160 630 calls);
- 44 • we show that automatically detected valence and anger are linked with CSAT scores and response rate;
- 45 • we qualify the selection bias inherent in self-reported CSAT by showing that positive valence and anger are  
46 respectively over- and under-represented in the collected scores;
- 47 • we show that the dynamics of emotion in each call is relevant with respect to both scores and response rate.

## 48 2 Related work

49 Natural language and speech processing techniques have been explored for the purpose automatic quality monitoring.  
50 Here we review related work pertaining to automatic emotion detection, customer satisfaction, or both.

### 51 2.1 Emotion recognition in phone calls

52 Typical call center conversations focus on at most a handful of problems or requests, and generally last a few minutes.  
53 There is a large body of work on emotion recognition showing that (non-neutral) emotional events are often rare, with  
54 typically only a few emotionally colored utterances per call. This results in highly unbalanced data, with a predominance  
55 of neutral labels [Morrison et al., 2007].

56 Since customer dissatisfaction can have a disproportional impact on a company's reputation and function, negative  
57 emotions have been a central focus of many works [Vaudable and Devillers, 2012, Erden and Arslan, 2011, Morrison  
58 et al., 2007]. In [Galani et al., 2013], a corpus of 135 call centre conversations was annotated using a fine-grained  
59 approach aiming at specific emotions such as pleasure, satisfaction, surprise, interest, anger, irritation, frustration,  
60 anxiety. These categorical values were then grouped under macro-classes such positive and negative. Vaudable and  
61 Devillers [2012] followed a similar approach with three macro-classes representing neutral (no emotion is expressed),  
62 negative (containing anger, disappointment, and negative-surprise), positive (satisfaction, positive-surprise).

63 Most studies highlight the complexity of detecting emotions in real-life phone calls. In addition to class imbalance,  
64 many authors note that emotions in call center interactions are more nuanced than what is found in databases of  
65 prototypical emotion [Vaudable and Devillers, 2012]. This complexity usually results in moderate inter-annotator  
66 agreement [Galani et al., 2013, Vaudable and Devillers, 2012] as well as lower prediction performance.

67 All reviewed methods rely on machine learning algorithms taking as input acoustic and/or linguistic features. In this  
68 study, both types of features are used. Most of the previous works on emotion recognition in phone calls focus on  
69 discrete emotions [Vaudable and Devillers, 2012, Erden and Arslan, 2011], mainly "anger". This focus is motivated  
70 by the fact that "anger" is among the most frequent negative emotions expressed by customers when calling call  
71 centers [Petrushin, 1999]. "Anger" is also among is usually found among best recognized by automatic approaches  
72 [Deschamps-Berger et al., 2021]. In addition, "anger" is considered to be the most important emotion for business  
73 [Petrushin, 1999]: detection of angry customers is central in the activity of companies, and several of them take

74 into account "anger" in their processes. For this purpose, we also consider "anger" in this paper. However, modern  
75 approaches of affective computing exploit dimensional representations that are also considered in this paper. More  
76 precisely, in this work emotion is modeled with two independent labels: an anger flag, and a "positive/negative valence"  
77 flags.

## 78 2.2 Prediction of customer satisfaction and/or service quality

79 Due to their optional nature and typically low response rates, CSAT surveys can miss important information. Automatic  
80 satisfaction predictions aims at filling the gaps by analyzing and rating all interactions. In Zweig et al. [2006], the  
81 authors design expert features from transcripts produced by Automatic Speech Recognition (ASR) to tag a call as a  
82 "good" or "bad" interaction. For a given amount of listening effort, this method triples the number of "bad" calls that  
83 are identified, over a policy of randomly sampling calls.

84 In contrast to approaches based mainly on expert features, the trend has more recently been to predict quality metrics  
85 with machine learning as a main tool. For instance Auguste et al. [2019] present classifiers aimed at labeling the  
86 customer in each call as "a promoter", "passive" or "a detractor". Ratings obtained from end-of-call questionnaires are  
87 used as the ground truth. In order to address the ordinal, subjective and skewed nature of self-reported satisfaction, a  
88 ranking approach is considered by Bockhorst et al. [2017]. The approach produces more accurate predictions compared  
89 to standard regression and classification approaches that directly fit the survey scores with call data.

90 Besides lexical and acoustic features, characteristics specific to dyadic interactions have been used to predict satisfaction.  
91 In [Chowdhury et al., 2016], turn-taking characteristics including participation equality, turn-taking freedom and  
92 statistics related speaker turns are used as input to a predictor. This approach is used to predict the final emotional  
93 manifestation of a conversation, which is considered as the satisfaction of the customer (positive, negative or neutral).  
94 The authors show that turn-taking features outperform lexical and prosodic feature sets. In [Luque et al., 2017], ASR  
95 transcripts, dialog turn-level features and acoustic/prosodic features are combined to predict customer satisfaction. The  
96 experimental results suggest that verbal communication convey more information than non-verbal cues with respect to  
97 customer satisfaction and that both sources of information are complementary. A Deep Convolutional Neural Network  
98 (CNN) is proposed to embed both linguistic and acoustic/prosodic features which allow to learn a representation able to  
99 capture customer's satisfaction.

100 The joint modeling of customer satisfaction at turn- and call-levels has been investigated in Ando et al. [2017, 2020].  
101 Two types of long short-term memory recurrent neural networks (LSTM-RNNs) are proposed to capture contextual  
102 information and the relationship between call-level and turn-level customer satisfaction. This hierarchical approach  
103 outperforms SVM based approach with relative error reductions of over 20%. These works show that the interplay  
104 between verbal and non-verbal communication for the prediction of customer satisfaction is complex but could be  
105 exploited to improve the performance of automated quality monitoring systems.

106 In [Segura et al., 2016], Deep CNNs are employed for feature learning for continuous prediction of satisfaction. The  
107 lower "feature representation" layers from a conflict detection model trained on TV media are used as feature extractors  
108 in a satisfaction prediction tasks. The authors demonstrate that the learned features overpower traditional spectral  
109 features, showing thereby a potential for domain transfer.

## 110 2.3 Joint analysis of emotion and customer satisfaction

111 In [Kim et al., 2020], the authors use detected sentiment in speech as features for predicting self-reported satisfaction,  
112 showing that this method enabled the system to predict CSAT nearly as well as a human listener. They also found  
113 that valence is the sentiment most linked to CSAT. In [Chowdhury et al., 2016], the authors show that user satisfaction  
114 could be modeled as a the final emotional manifestation (positive, negative and neutral) of a conversation. In Luque  
115 et al. [2017], the authors analyzed the significance of various acoustic, prosodic and linguistic features that correlate  
116 to emotion to predict self-reported satisfaction in contact centre phone calls. Linguistic features are obtained by a  
117 customized Automatic Speech Recognition (ASR) system trained with calls from call centers. Fundamental frequency,  
118 speech loudness as well articulation rate (number of syllable nuclei per phonation time) are the main acoustic and  
119 prosodic features used in this work. The authors also consider Low Level Descriptors extracted with OpenSmile using  
120 paralinguistic 2013 configuration [Schuller et al., 2013]. We exploit a similar methodology in this paper by combining a  
121 customized ASR system (section 5.1) and low-level descriptors for emotion recognition (section 5.2).

122 As noted above, some works use emotions as a proxy for satisfaction, while other works use them as inputs to supervised  
123 predictors trained on self-reported satisfaction. By contrast, this study is primarily descriptive. Our goal is to better  
124 understand the relationship between emotion and self-reported satisfaction. More precisely, we ask whether customers  
125 expressing positive/negative emotions respond more to CSAT questionnaires and report higher/lower satisfaction.

## 126 3 Hypotheses

127 The aim of this study is to determine if and how detected emotions and self-reported CSAT are related. We consider  
 128 both aggregate (call-level) and dynamic indicators. Since filling the CSAT questionnaire is optional, we also consider  
 129 the response rate, defined as the proportion of calls with a CSAT score over all calls.

130 We formulate the following hypotheses:

### 131 • H1 Customers' emotions and CSAT response

- 132 – H1a: Customers expressing positive emotions respond more to CSAT questionnaires;
- 133 – H1b: Customers expressing negative emotions respond less to CSAT questionnaires;
- 134 – H1c: Customers expressing anger respond less to CSAT questionnaires;

### 135 • H2 Customers' emotions and self-reported satisfaction

- 136 – H2a: Customers expressing positive emotions report higher satisfaction;
- 137 – H2b: Customers expressing negative emotions report lower satisfaction;
- 138 – H2c: Customers expressing anger report lower satisfaction;

### 139 • H3 Customers' emotional profiles and CSAT response rate

- 140 – H3a: Customers manifesting upward positive valence dynamics (more positive emotions towards the end  
 141 of the call) exhibit a higher CSAT response rate compared to flat or negative dynamics;
- 142 – H3b: Customers manifesting downward negative valence dynamics (fewer negative emotions towards the  
 143 end of the call) exhibit a higher CSAT response rate compared to flat or positive dynamics;
- 144 – H3c: Customers manifesting downward anger dynamics (fewer anger events towards the end of the call)  
 145 exhibit a higher CSAT response rate;

### 146 • H4 Customers' emotional profiles and self-reported satisfaction

- 147 – H4a: Customers manifesting upward positive valence dynamics report higher satisfaction;
- 148 – H4b: Customers manifesting downward negative valence dynamics report higher satisfaction;
- 149 – H4c: Customers manifesting downward anger dynamics report higher satisfaction;

## 150 4 Materials

### 151 4.1 Database

152 The corpus consists of 160 630 call center conversations that occur between a customer and an operator in French. All  
 153 conversations were recorded between July 2021 and September 2021. The corpus contains a total of 28 478 hours of  
 154 conversation, with call duration ranging from 40 seconds to 80 minutes. The average call duration is 11 minutes. All  
 155 calls were recorded in stereo, with a sample rate 8 kHz and 16 bit format.

156 Calls were automatically processed and analyzed using custom models. The linguistic content of the calls was extracted  
 157 using an Automatic Speech Recognition (ASR) system tuned on phone conversations (section 5.1). Personal information  
 158 such as names, addresses and phone numbers was detected and permanently removed using a Named Entity Recognition  
 159 (NER) algorithm.

### 160 4.2 Customer satisfaction

161 Operators may but do not always propose the CSAT questionnaire at the end of each call, and the customer is free to  
 162 accept or decline the eventual proposal. In this context, it is useful to define the response rate as the number of calls  
 163 with a CSAT score over the total number:

$$\text{CSAT response rate} = \frac{\text{Nbr of Calls with CSAT scoring}}{\text{Total Nbr of Calls}} \quad (1)$$

164 The CSAT questionnaire takes the form of a 9-point Likert: from level 0 (very unsatisfied) to 9 (very satisfied). The  
 165 value obtained for each call is the CSAT score ( $CSAT_{score}$ ).

166 The results of the CSAT questionnaire are also studied using a binary coding. Customers who answer from 6 to 9 are  
 167 considered to be moderately to highly satisfied (labeled as "*High Satisfaction*"), those who answer from 0 to 5 are  
 168 labeled as dissatisfied customers (labeled as "*Low Satisfaction*").

169 In addition to the per-call raw CSAT score, we introduce the Aggregate High Satisfaction score (aggH-CSAT). Given  
 170 all calls with a CSAT score, we define aggH-CSAT as the number of calls with *High satisfaction* divided by the total  
 171 number of calls with a CSAT score (*Low and High satisfaction scores*):

$$\text{aggH-CSAT} = \frac{\sum_{i=6}^9 \text{CSATscore}(i)}{\sum_{i=1}^9 \text{CSATscore}(i)} \quad (2)$$

172 The aggH-CSAT score is suitable for measuring the proportion of high self-reported customer satisfaction in calls with  
 173 and without a detected emotion. aggH-CSAT score is bounded in  $[0, 1]$ , a greater value indicating a greater proportion  
 174 of higher satisfaction.

### 175 4.3 Affective Indicators

176 We define positive emotional events as utterances flagged as positive/negative by the valence classifier (nega-  
 177 tive/neutral/positive). Similarly, anger events are utterances flagged as containing anger by the anger classifier. An  
 178 utterance is a speaking turn containing at least three words. Turns range from 3s to 30s. A call is considered to contain  
 179 the emotion (anger, positive.negative valence) if one emotion event is detected (hard labels). The satisfaction-emotion  
 180 relationship is studied independently for each emotion.

181 For each emotion, we compute the CSAT response rate based on the number of calls with and without CSAT scoring.  
 182 We compare the CSAT response rate of calls of customers expressing specific emotions. We then compare the proportion  
 183 of calls ( $\pi_1$  and  $\pi_2$ ) of customers expressing an emotion (i) with and without CSAT scoring and (ii) with *High and Low*  
 184 *satisfaction* scores.

### 185 4.4 Dynamics of Affective Indicators and Customers' profiles

186 We analyze the dynamics of affective indicators to better characterize their relationship with CSAT. Given that call  
 187 duration varies, we define three call phases using interquartile ranges:

- 188 • beginning phase, defined as first 25% of the duration (lower quartile).
- 189 • middle phase, defined as the next 50% of the call (middle quartile).
- 190 • end phase, defined as the last 25% of the call (upper quartile). The CSAT questionnaire is typically proposed  
 191 during the conclusion of the call, thus in the end phase.

192 To study the dynamics of emotions during the call, the affective indicators are extracted for each phase of the call, for  
 193 each emotion:  $N_{beg}, N_{mid}, N_{end}$  denoting the number of emotional events detected in the phase. We define the  $\Delta$   
 194 score, a summary metric of the dynamics of detected emotions:

$$\Delta = \frac{N_{end} - N_{beg}}{N_{beg} + N_{mid} + N_{end}} \quad (3)$$

195 A negative/positive  $\Delta$  score indicates a decrease/increase in the occurrence of emotional events as the call progresses.  
 196  $\Delta$  score is used to identify profiles of customers manifesting upward / downward emotion dynamics.

## 197 5 Methods

198 This section presents our methods for detecting affective indicators and analyzing their relationship with self-reported  
 199 satisfaction. An essential step of our process is automatic speech recognition, which is described in section 5.1. We  
 200 then describe the automatic emotion recognition model in section 5.2.

### 201 5.1 Automatic speech recognition

202 The ASR system used for this study draws from the Eesen framework proposed Miao et al. [2015], and is implemented  
 203 using a combination of custom code and utilities from Kaldi [Povey et al., 2011] and Eesen [Miao et al., 2015].



204 The features are Mel Frequency Cepstral Coefficients (MFCCs) computed with their deltas and delta-deltas over  
205 20ms windows using 20 Mel-frequency bins, resulting in one 60-dimensional vector per frame. Cepstral Mean and  
206 Variance Normalization (CMVN) is applied using training data statistics to obtain identical per-utterance statistics  
207 across utterances, a technique known to improve the robustness of ASR systems to acoustic variability [Viikki and  
208 Laurila, 1998].

209 The phonetic recognizer is a bi-directional LSTM network with 5 layers, with a cell dimension of 320. It is trained using  
210 CTC loss [Graves et al., 2006], whose main benefit is to not require alignment of the phonemes in the training data.

211 The language model computes probabilities of  $n$ -grams, smoothed using the Kneser-Ney method to estimate the  
212 probabilities of  $n$ -grams unseen in the training data. We compute smoothed 1-, 2- and 3-grams probabilities.

213 The decoding graph, a weighted Forward State Transducer (wFST), incorporates the list of phonemes, the phonetic  
214 lexicon associating each word with its pronunciation(s), and the language model. During decoding, the phonetic model  
215 outputs phoneme probabilities for each frame, which are passed as inputs to the decoding graph. The most probable  
216 word sequence is computed using Viterbi decoding.

217 The hyper-parameters of the ASR system (for feature extraction, phonetic modeling, language modeling and decoding)  
218 were chosen using results of previous optimization experiments in similar domains.

219 Separate ASR models were trained for the agent and customer channels. The training data consists of 104 hours  
220 collected from a similar domain and manually transcribed. 5 hours are set aside for evaluation, leaving 99 hours of  
221 training data. Using the method described above (previously optimized in similar contexts), the ASR system achieves  
222 a Word Error Rate (WER) of 25.5% on the customer channel, and 16.5% on the agent channel. Only the customer  
223 channel ASR model is used in this study.

## 224 5.2 Automatic emotion recognition

### 225 5.2.1 Data and labels

226 Emotions in the customer’s speech were detected using pre-trained custom models. Training data originated from  
227 call center data and was annotated by internal annotators using our recently open-sourced speech annotation platform  
228 Labelit.<sup>1</sup>

229 Annotators were asked to label utterances extracted from call center data. Valence was annotated on an ordinal scale  
230 using a 5-point Likert scale, with multiple annotators (3 to 5) annotating each utterance. Our operational definition  
231 of valence is directly taken from the dimensional valence-arousal model [Russell, 1980]. Ordinal labels were binned  
232 to make up three categories: negative (1-2), neutral (3) and positive valence (4-5). Categorical training labels were  
233 obtained by averaging the ordinal labels of the multiple annotators. Then, we match the values with the corresponding  
234 bin (1-2; 3; 4-5). Anger was annotated as a binary categorical label, with multiple annotators (3 to 5) annotating each  
235 utterance. Categorical training labels were obtained by majority vote at the utterance level.

### 236 5.2.2 Emotion Recognition model

237 Multiple model architectures and feature extraction methods were optimized and compared on this task. We now  
238 describe the best performing pipeline on this dataset.

239 We designed two predictive models one for respectively valence (3-class problem) and anger (2-class problem) prediction  
240 that exploit both linguistics and non-linguistics features.

241 The custom ASR model described in section 5.1 was applied on each utterance in the analyzed dataset. We extracted  
242  $n$ -grams, with  $n$  ranging from 1 to 4. The textual content from each utterance is then vectorized using Term Frequency  
243 Inverse Document Frequency (TD-IDF).

244 Summative audio features were extracted for each utterance audio, resulting in a fixed-sized vector containing a  
245 combination of low-level descriptors (LLDs) and associated functionals. This technique is commonly used in the  
246 literature [Schuller et al., 2010a, 2013, 2019]. In this paper, the openSMILE toolkit [Eyben et al., 2013] was used to  
247 extract the INTERSPEECH 2010 Paralinguistic Challenge [Schuller et al., 2010b] feature set from the speech signal.  
248 The set contains 1582 features for each utterance. OpenSMILE computes Low-Level Descriptors (LLDs) from pitch,  
249 loudness, voice quality as well as cepstrum and linear predictive representations. Then, a series of functionals are  
250 applied to LLDs such as extremes, statistical moments, percentiles, duration and regression.

---

<sup>1</sup><https://github.com/voicelab-org/labelit/>

251 We optimized and evaluated multiple models for each task (anger, valence). Separate logistic regression models were  
 252 trained for audio and text features using a 5-fold cross-validation strategy. The final prediction is a weighted sum of the  
 253 best audio and text models, where the weights are learnt. See Table 1 for a performance summary.

Table 1: Emotion recognition system: performances

	Valence prediction	Anger prediction
$F1$ (weighted)	0.67	0.69

### 254 5.3 Data analysis

255 In order to test the hypotheses formulated in 3, we consider the following variables:

- 256 • CSAT response rate (equation 1): variable between 0 to 1, defined as the number of calls with a customer  
 257 scoring divided by the total number of calls.
- 258 • Aggregate satisfaction score (aggCSAT, equation 2): variable between 0 to 1, defined as the number of calls  
 259 with *High Satisfaction* (i.e., CSAT score [6, 9]) divided by the total numbers of a calls with a CSAT score.

260 A chi-square test was used to compare CSAT response and High/Low satisfaction against detected emotion (**H1**, **H2**  
 261 and **H3** and **H4**) using a significance level of  $\alpha = 0.05$ , in case of multiple comparisons a Bonferroni correction has  
 262 been performed. Analysis results were presented as frequency for categorical data.

263 Using a two proportion z-test, we analyze proportions of emotions in calls with and without CSAT scoring (**H1a**, **H1b**  
 264 and **H1c**). For each analysis, we compute the two sample proportions  $\pi_1$  and  $\pi_2$ . We then use the following null  
 265 hypothesis:

- 266 •  $H_0$ :  $\pi_1 = \pi_2$  (the two sample proportions are equal)

267 We consider the corresponding alternative hypotheses that can be either left-tailed, or right-tailed:

- 268 •  $H_a$  (left-tailed):  $\pi_1 < \pi_2$  (sample 1 proportion is less than sample 2 proportion)
- 269 •  $H_a$  (right-tailed):  $\pi_1 > \pi_2$  (sample 1 proportion is greater than sample 2 proportion)

270 using a significance level of  $\alpha = 0.05$  to indicate strong evidence against the null hypothesis  $H_0$ . Bonferroni corrections  
 271 were performed when multiple comparisons are made.

272 For each analysis, we report the proportion, the z-statistic and the significance level.

## 273 6 Results

### 274 6.1 CSAT scoring in phone calls

275 In this section, we analyze (i) the CSAT response rate; and for calls with a scoring, (ii) the distribution of the CSAT  
 276 scores. The distribution detailed in (Table 2) shows that most of the customers do not respond to the satisfaction  
 277 questionnaire with a CSAT response rate of approximately of 30%.

Table 2: Distribution of call types: with and without customer satisfaction scoring (CSAT)

	# of calls
All calls	160 630
Without CSAT scoring	111 001
With CSAT scoring	49 629
CSAT response rate	0.29
Low satisfaction (CSAT score 0-5)	3 120
High satisfaction (CSAT score (6-9)	46 309
aggH-CSAT	0.93

278 However, when customers do respond to the questionnaire, they tend to report high satisfaction as reported in table 2  
 279 and figure 1. The aggregate high satisfaction score (aggH-CSAT) is very high 0.93.



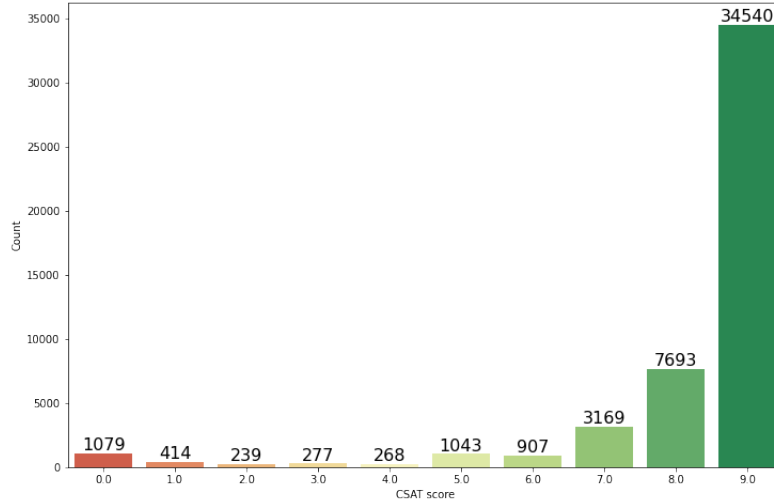


Figure 1: CSAT scoring distribution.

280 Since CSAT is collected on a fraction of all calls, the CSAT scores do not provide a full view of customer satisfaction,  
 281 as the population of non-responders may differ greatly from the population of responders.

## 282 6.2 Emotion recognition

283 In this section, we apply the two emotion prediction models described in section 5.2 to label 160 630 calls of the  
 284 database of the current study (table 2). As mentioned in section 4.3, each emotion target is studied independently.

285 The number of calls detected per emotion by our models are reported in table 3. Consistent with the literature [Vaudable  
 286 and Devillers, 2012, Morrison et al., 2007], calls containing anger are somewhat rare.

Table 3: Number of calls per detected emotion

# of calls	Positive valence	Negative Valence	Anger
Target class	40 703	99 629	4 478
No target class	119 927	61 001	156 152

## 287 6.3 CSAT response rate and emotions

288 In this section, we compare the CSAT response rate against detected emotional events (positive/negative valence and  
 289 anger) using the data described in table 3. In table 4 and figure 2, we report the number of calls with and without  
 290 a CSAT score per emotional category. We also report the number of calls with and without CSAT scoring without  
 291 emotional analysis ("All calls") and we take the CSAT response of such calls as our baseline. (0.29).

Table 4: CSAT response rate and emotion in respect to the emotion analysis. \* indicates a significant difference with the baseline ( $p < 0.05$ )

# of calls	With CSAT scoring	Without CSAT Scoring	CSAT response rate
All calls	49 629	111 001	0.29
Calls with positive valence	13 491	27 212	<b>0.33*</b>
Calls with negative valence	31 722	67 907	<b>0.32*</b>
Calls with anger	892	3 586	<b>0.20*</b>

292 Table 4 shows that the proportions of calls with a CSAT scoring and with positive valence differ from "All Calls"  
 293 ( $\chi(1) = 76.18, p < 0.05$ ). A higher CSAT response rate is observed (0.33) for calls with positive valence. The  
 294 analysis indicates similar results regarding calls with negative valence ( $\chi(1) = 25.43, p < 0.05$ ) with a CSAT response  
 295 rate of 0.32. A Chi-Square test was performed to determine whether the proportion of calls with CSAT scoring was

296 equal between "All Calls" and "Calls with Anger". The proportions did differ ( $\chi(1) = 259.43, p < 0.05$ ). Customers  
 297 expressing anger significantly respond less to the CSAT questionnaire (0.20).

298 We now analyze the proportion of detected emotions in calls with and without CSAT scoring using a z-test proportion  
 299 test (table 5). The results indicate that the proportion of calls with positive valence is higher in calls with CSAT scoring  
 300 (with CSAT 0.27 vs. without CSAT 0.24,  $z = 11.36, p < 0.05$ ). A similar result is obtained for calls with negative  
 301 valence (with CSAT 0.63 vs. without CSAT 0.61,  $z = 10.46, p < 0.05$ ). However, the proportions of calls with anger  
 302 is significantly lower in calls with a CSAT scoring ( $0.01 < 0.03, z = -16.12, p < 0.05$ ).

Table 5: Proportion of detected emotion in calls with and without CSAT scoring.

Proportion in	With CSAT scoring ( $\pi_1$ )	Without CSAT Scoring ( $\pi_2$ )	$\pi_1$ vs. $\pi_2$
Calls with positive valence	0.27	0.24	$z = 11.36, p < 0.05$
Calls with negative valence	0.63	0.61	$z = 10.46, p < 0.05$
Calls with anger	0.01	0.03	$z = -16.12, p < 0.05$

### 303 6.4 Satisfaction score and emotions

304 When CSAT scoring is available, we analyze the distribution of CSAT scores with respect to the presence of emotional  
 305 content. In figure, 3, we report the normalized CSAT score distribution (0-9). Call volumes for each analysis are reported  
 306 in table 6.

Table 6: Number of calls per CSAT scoring in respect of affective computing analysis (see also figure 3) and the Aggregate High Customer Satisfaction score (aggH-CSAT) in respect to the emotion analysis. \* indicates a significant difference with the baseline ( $p < 0.05$ )

	0	1	2	3	4	5	6	7	8	9	aggH-CSAT
All calls	1079	414	239	277	268	1043	907	3169	7693	34540	0.93
Calls with positive valence	150	94	44	43	53	182	165	602	1814	10344	<b>0.96*</b>
Calls without positive valence	929	320	195	234	215	861	742	2567	5879	24196	<b>0.92*</b>
Calls with negative valence	883	299	180	184	193	747	556	2012	4819	21849	<b>0.92*</b>
Calls without negative valence	196	115	59	93	75	296	351	1157	2874	12691	<b>0.95*</b>
Calls with anger	184	17	17	11	10	39	26	51	115	422	<b>0.69*</b>
Calls without anger	895	397	222	266	258	1004	881	3118	7578	34118	<b>0.94*</b>

307 We consider as a baseline the CSAT score distribution of calls irrespective of detected emotional content ("*All calls*").  
 308 As we can see in figure 3 most of the customers report a high satisfaction. This is also supported by the aggregate high  
 309 satisfaction score (aggH-CSAT) (section 4.2), which captures a normalized balance between High/Low scoring, a high  
 310 score indicating more satisfactory customers. The aggH-CSAT score of calls is already high (0.93), which, as already  
 311 noted, shows that most of the customers report high CSAT scores.

312 To study the relationship between reported satisfaction and detected customer emotion, we compare the proportions  
 313 of High/Low satisfaction against the presence or absence of each emotion. (table 6 and figure 3). The proportions  
 314 of High satisfaction is higher for customers expressing positive valence than those not expressing it ( $\chi(1) = 184.1$ ,  
 315  $p < 0.05$ ). In addition, a higher aggH-CSAT is observed for such customers ( $0.96 > 0.93$ ). Regarding customers  
 316 expressing negative valence, similar results are obtained ( $\chi(1) = 184.85, p < 0.05$ ). A slightly lower aggH-CSAT is  
 317 observed ( $0.92 < 0.93$ ).

318 Customers expressing anger report much lower satisfaction (aggH-CSAT = 0.69). Customers expressing anger still  
 319 report high satisfaction on average, but a significant proportion reports dissatisfaction. (table 6 and figure 3). A  
 320 significantly higher satisfaction is observed for customers who are not expressing anger (aggH-CSAT = 0.94). The  
 321 proportions of high satisfaction differ in calls with and without anger ( $\chi(1) = 867.77, p < 0.05$ ).

### 322 6.5 CSAT response rate and emotional dynamics profiles

323 Using the  $\Delta$  score of calls of each emotion category (section 4.4), we categorized calls into upward and downward  
 324 emotional profiles according to the dynamics of detected emotion. For each emotion, only calls containing at least one  
 325 detected event are considered. In table 7, we report the CSAT response rates of such emotional profiles. The results  
 326 show that customers with upward anger dynamics (i.e., more anger events at the end of the call) respond significantly  
 327 less to the satisfaction score ( $0.16, p < 0.05$ ).

Table 7: CSAT response rate with respect to the customers' emotional profiles. \* indicates a significant difference with the baseline ( $p < 0.05$ )

# of calls	With CSAT scoring	Without CSAT Scoring	CSAT response rate
All calls	49 629	111 001	0.29
Calls with upward positive valence dynamics	11 896	24 577	<b>0.32*</b>
Calls with downward positive valence dynamics	302	521	<b>0.36*</b>
Calls with upward negative valence dynamics	15 489	40 017	<b>0.27*</b>
Calls with downward negative valence dynamics	5 854	9 621	<b>0.37*</b>
Calls with upward anger dynamics	437	2 155	<b>0.16*</b>
Calls with downward anger dynamics	82	230	<b>0.38*</b>

328 We compare the proportions of calls with CSAT scoring across profiles. We first compare the proportions of such profiles  
 329 to "All Calls". We observe significant differences with calls with upward positive valence dynamics ( $\chi(1) = 40.84$ ,  
 330  $p < 0.05$ ), downward positive valence dynamics ( $\chi(1) = 12.61$ ,  $p < 0.05$ ), with upward negative valence dynamics  
 331 ( $\chi(1) = 175.20$ ,  $p < 0.05$ ), downward negative valence dynamics ( $\chi(1) = 314.01$ ,  $p < 0.05$ ), upward anger dynamics  
 332 ( $\chi(1) = 235.69$ ,  $p < 0.05$ ) and downward anger dynamics ( $\chi(1) = 16.20$ ,  $p < 0.05$ )

333 We now compare the proportions of calls with and without CSAT in upward/downward emotional profiles. The  
 334 proportions of high satisfaction are different in calls with upward/downward positive valence dynamics ( $\chi(1) = 5.9$ ,  
 335  $p < 0.05$ ). Similar observations are made for calls with upward/downward positive negative dynamics ( $\chi(1) = 566.27$ ,  
 336  $p < 0.05$ ) and calls with upward/downward anger valence dynamics ( $\chi(1) = 16.2$ ,  $p < 0.05$ ).

### 337 6.6 Self-reported satisfaction and emotional dynamics profiles

338 Using the approach described in section 6.4, we compare the proportions of high satisfaction across dynamics profiles.

Table 8: Number of calls per CSAT scoring with respect to affective computing analysis (see also figure 3) and the Aggregate High Customer Satisfaction score (aggH-CSAT) in respect to the emotion analysis. \* indicates a significant difference with the baseline ( $p < 0.05$ ) (To be changed to integrating  $\Delta$  score)

	0	1	2	3	4	5	6	7	8	9	aggH-CSAT
All calls	1079	414	239	277	268	1043	907	3169	7693	34540	0.93
Calls with upward positive valence dynamics	110	84	38	38	45	156	142	542	1582	9159	<b>0.96*</b>
Calls with downward positive valence dynamics	5	0	0	0	2	6	5	12	40	232	<b>0.95*</b>
Calls with upward negative valence dynamics	678	189	123	117	124	451	287	1059	2366	10095	<b>0.89*</b>
Calls with downward negative valence dynamics	89	36	20	21	23	102	76	326	880	4281	<b>0.95*</b>
Calls with upward anger dynamics	135	11	12	7	8	20	9	18	42	175	<b>0.55*</b>
Calls with downward anger dynamics	5	1	1	0	0	6	1	7	15	46	<b>0.84*</b>

## 339 7 Discussion

340 It is natural to expect that positive/negative emotions are associated with higher/lower satisfaction, and the results  
 341 presented above mostly confirm this expectation.

342 More precisely, we found that customers expressing positive emotions respond more to CSAT questionnaire (**H1a**:  
 343 CSAT response rate 0.33) and (when they respond) report higher satisfaction (**H2a**: aggH-CSAT = 0.96). **All null**  
 344 **hypotheses regarding anger events were rejected.** Customers expressing anger respond significantly less to the CSAT  
 345 questionnaire (**H1c**: CSAT response rate=0.2) and report lower satisfaction on average (**H2c**: aggH-CSAT = 0.69).

346 We also found that the proportions of positive emotions in calls with CSAT scoring is higher than in calls without  
347 CSAT scoring ( $0.27 > 0.24$ ,  $z = 11.36$ ,  $p < 0.05$ ). An opposite observation is made for calls with anger ( $0.01 < 0.03$ ,  
348  $z = -16.12$ ,  $p < 0.05$ ).

349 Contrary to our hypothesis **H1b**, customers expressing negative emotions do not respond less to CSAT questionnaires  
350 (CSAT response rate: 0.32) but we do observe slightly lower overall satisfaction ( $\text{aggH-CSAT} = 0.92$ ). The weakness  
351 of the correlation here suggests that the presence of negatively valent emotion alone has a small impact on satisfaction.  
352 Negative valence covers a wide swath of emotions which can have many different causes, including the customer's  
353 mood, the quality of the interaction with the company representative, issues and obstacles for which the company or the  
354 representative may or may not be responsible. For example, a customer may call to express frustration about a problem,  
355 then report high satisfaction after the problem was well handled.

356 These results suggest that automatic emotion recognition by itself can complement but not replace self-reported  
357 satisfaction. While negative/positive emotions are linked with lower/higher CSAT, the presence of an emotion alone  
358 (e.g. anger) does not fully account for the reported score. Satisfaction is influenced by many factors besides momentary  
359 emotion, such as response effectiveness, overall service quality and other interactions with the company prior to the call.  
360 Furthermore, due to the limitations of optional reporting, we cannot be certain that the observed correlations apply in  
361 precisely the same way for calls for which no score is given.

362 The analysis of the response rate with respect to detected emotion provides insight into the selection bias induced by  
363 optional self-reporting of satisfaction. Angry customers have a tendency to answer less, and "happy" customers have a  
364 tendency to answer more, to CSAT questionnaires. Given this, it is likely that the aggregate CSAT score gives an overly  
365 optimistic picture of customer satisfaction. The reasons for the observed difference in response rates are unclear. As  
366 noted previously, proposal of the CSAT questionnaires is left to the discretion of customer representatives. It is possible  
367 that the emotions expressed by the customer not only influence whether the customer will respond, but also whether the  
368 questionnaire is offered in the first place.

369 When we compare the dynamics of detected anger against CSAT scores and response rate, we observe significant  
370 differences between "upward" and "downward" profiles. Customers manifesting downward anger dynamics exhibit a  
371 higher CSAT response rate (0.38, **H3c**) and satisfaction ( $\text{aggH-CSAT} = 0.84$ , **H4c**) than customers manifesting upward  
372 anger dynamics (CSAT response rate=0.16 and  $\text{aggH-CSAT} = 0.55$ ). This suggests that anger expressed towards the  
373 end of the call is more meaningful with respect to satisfaction.

374 Customers manifesting upward positive valence dynamics exhibit a higher than average CSAT response rate (0.32,  
375 **H3a**) and satisfaction ( $\text{aggH-CSAT} = 0.96$ , **H4a**). However, we observed a similar behaviour for calls with downward  
376 positive valence dynamics (CSAT response rate=0.36 and  $\text{aggH-CSAT} = 0.95$ ). Downward positive valence dynamics  
377 are actually associated with a *higher* response rate than upward dynamics. The observations are not symmetrical to  
378 those regarding anger. This result shows that positive valence is a relevant indicator of customer satisfaction since  
379 customers manifesting both upward and downward dynamics exhibit a high CSAT response rate and satisfaction.

380 Consistent with our findings for anger, calls with upward negative valence dynamics have a lower CSAT response rate  
381 (0.27) and reported high satisfaction ( $\text{aggH-CSAT} = 0.89$ ) compared to calls with downward negative valence (CSAT  
382 response rate=0.37 and  $\text{aggH-CSAT} = 0.95$ , **H3b** and **H4b**).

## 383 8 Conclusion and future works

384 This work provides a detailed description of the relationship between automatically detected emotion and self-reported  
385 satisfaction and shows that, for the studied dataset, detected valence and anger are linked with CSAT scores; positive  
386 emotions are linked with higher response rates, while anger is linked with a lower response rate; and finally, the  
387 dynamics of emotion significantly weigh on both scoring and response rate.

388 These findings suggest that emotions could be used by companies as a complement to CSAT, especially to shed light on  
389 calls without CSAT scores. Automatically detected emotions could also be used as input features to CSAT predictors,  
390 keeping in the mind that if the goal is to automate CSAT, both the presence of the score and the score itself should be  
391 modeled.

392 Future work will include expanding the range of considered emotions, specifically including more dimensional (arousal,  
393 dominance) and categorical (surprise, joy, disgust, etc.) labels. This could take the form of probabilistic class labels to  
394 compute emotional profiles [Mower et al., 2011].

395 We also plan on developing CSAT predictors that jointly model response and scoring using the insights garnered in the  
396 present study.

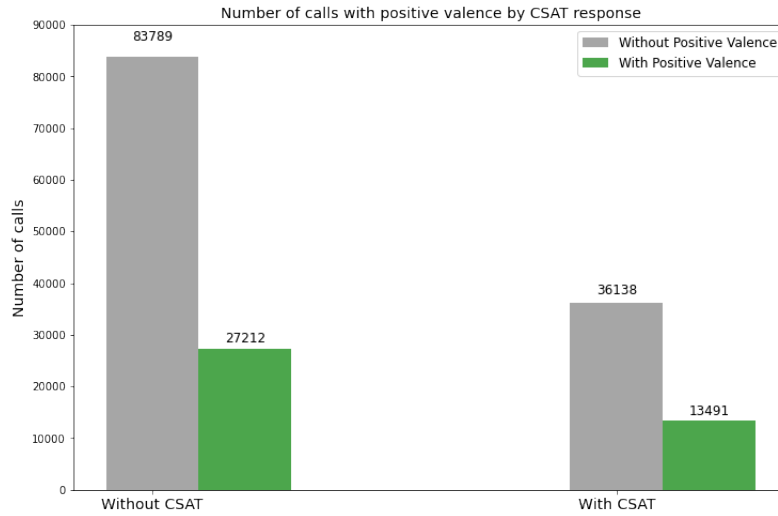
397 **References**

- 398 A. Ando, R. Masumura, H. Kamiyama, S. Kobashikawa, and Y. Aono. Hierarchical LSTMs with Joint Learning for  
399 Estimating Customer Satisfaction from Contact Center Calls. In *Proc. Interspeech 2017*, pages 1716–1720, 2017.  
400 doi:10.21437/Interspeech.2017-725.
- 401 A. Ando, R. Masumura, H. Kamiyama, S. Kobashikawa, Y. Aono, and T. Toda. Customer satisfaction estimation  
402 in contact center calls based on a hierarchical multi-task model. *IEEE/ACM Transactions on Audio, Speech, and*  
403 *Language Processing*, 28:715–728, 2020. doi:10.1109/TASLP.2020.2966857.
- 404 J. Auguste, D. Charlet, G. Damnati, F. Bechet, and B. Favre. Can we predict self-reported customer satisfaction from  
405 interactions? In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*  
406 *(ICASSP)*, pages 7385–7389, 2019. doi:10.1109/ICASSP.2019.8683896.
- 407 J. Bockhorst, S. Yu, L. Polania, and G. Fung. Predicting self-reported customer satisfaction of interactions with a  
408 corporate call center. In Y. Altun, K. Das, T. Mielikäinen, D. Malerba, J. Stefanowski, J. Read, M. Žitnik, M. Ceci,  
409 and S. Džeroski, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 179–190, Cham, 2017.  
410 Springer International Publishing. ISBN 978-3-319-71273-4.
- 411 S. A. Chowdhury, E. A. Stepanov, and G. Riccardi. Predicting user satisfaction from turn-taking in spoken conversations.  
412 In *INTERSPEECH*, 2016.
- 413 T. Deschamps-Berger, L. Lamel, and L. Devillers. End-to-End Speech Emotion Recognition: Challenges of Real-Life  
414 Emergency Call Centers Data Recordings. In *2021 9th International Conference on Affective Computing and*  
415 *Intelligent Interaction (ACII)*, Nara, Japan, Sept. 2021.
- 416 M. Erden and L. M. Arslan. Automatic detection of anger in human-human call center dialogs. In *Proc. Interspeech*  
417 *2011*, pages 81–84, 2011. doi:10.21437/Interspeech.2011-21.
- 418 F. Eyben, F. Weninger, F. Gross, and B. Schuller. Recent developments in opensmile, the munich open-source  
419 multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM  
420 '13, page 835–838, New York, NY, USA, 2013. Association for Computing Machinery. ISBN 9781450324045.  
421 doi:10.1145/2502081.2502224. URL <https://doi.org/10.1145/2502081.2502224>.
- 422 D. Galanis, S. Karabetsos, M. Koutsombogera, H. Papageorgiou, A. Esposito, and M.-T. Riviello. Classification  
423 of emotional speech units in call centre interactions. In *2013 IEEE 4th International Conference on Cognitive*  
424 *Infocommunications (CogInfoCom)*, pages 403–406, 2013. doi:10.1109/CogInfoCom.2013.6719279.
- 425 A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber. Connectionist temporal classification: Labelling unsegmented  
426 sequence data with recurrent neural networks. In *Proceedings of the 23rd International Conference on Machine*  
427 *Learning, ICML '06*, page 369–376, New York, NY, USA, 2006. Association for Computing Machinery. ISBN  
428 1595933832. doi:10.1145/1143844.1143891.
- 429 Y. Kim, J. Levy, and Y. Liu. Speech sentiment and customer satisfaction estimation in socialbot conversations. In  
430 *INTERSPEECH*, 2020.
- 431 J. Luque, C. Segura, A. Sánchez, M. Umbert, and L. A. Galindo. The Role of Linguistic and Prosodic Cues on the  
432 Prediction of Self-Reported Satisfaction in Contact Centre Phone Calls. In *Proc. Interspeech 2017*, pages 2346–2350,  
433 2017. doi:10.21437/Interspeech.2017-424.
- 434 Y. Miao, M. Gowayyed, and F. Metze. Eesen: End-to-end speech recognition using deep rnn models and wfst-based  
435 decoding. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 167–174,  
436 2015. doi:10.1109/ASRU.2015.7404790.
- 437 D. Morrison, R. Wang, and L. C. De Silva. Ensemble methods for spoken emotion recognition in call-centres. *Speech*  
438 *Communication*, 49(2):98–112, 2007. ISSN 0167-6393. doi:<https://doi.org/10.1016/j.specom.2006.11.004>.
- 439 E. Mower, M. J. Matarić, and S. Narayanan. A framework for automatic human emotion classification using  
440 emotion profiles. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(5):1057–1070, 2011.  
441 doi:10.1109/TASL.2010.2076804.
- 442 V. A. Petrushin. Emotion in speech: Recognition and application to call centers. 1999.
- 443 D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz,  
444 J. Silovsky, G. Stemmer, and K. Vesely. The kaldi speech recognition toolkit. In *IEEE 2011 Workshop on Automatic*  
445 *Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011. IEEE Catalog No.: CFP11SRW-  
446 USB.
- 447 J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980.
- 448 B. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. A. Müller, and S. S. Narayanan. The interspeech 2010  
449 paralinguistic challenge. In *INTERSPEECH*, 2010a.

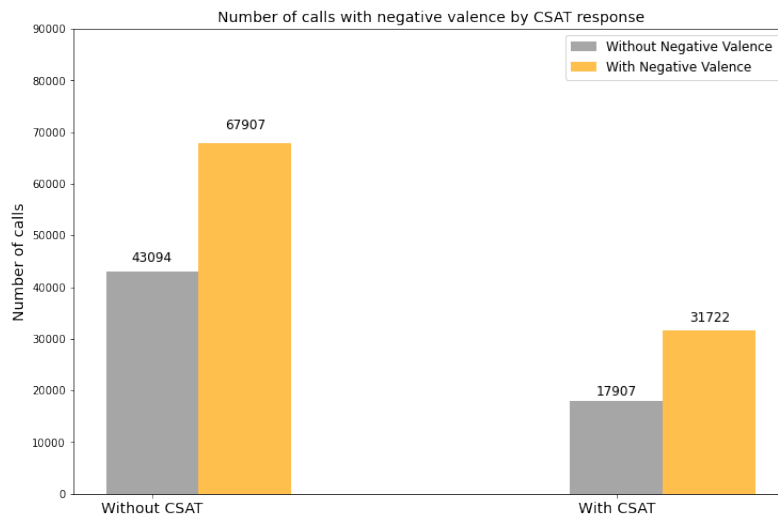


- 450 B. Schuller, F. Weninger, Y. Zhang, F. Ringeval, A. Batliner, S. Steidl, F. Eyben, E. Marchi, A. Vinciarelli, K. Scherer,  
451 M. Chetouani, and M. Mortillaro. Affective and behavioural computing: Lessons learnt from the first com-  
452 putational paralinguistics challenge. *Computer Speech & Language*, 53:156–180, 2019. ISSN 0885-2308.  
453 doi:<https://doi.org/10.1016/j.csl.2018.02.004>.
- 454 B. W. Schuller, S. Steidl, A. Batliner, F. Burkhardt, L. Devillers, C. A. Müller, and S. S. Narayanan. The INTERSPEECH  
455 2010 paralinguistic challenge. In *INTERSPEECH 2010, 11th Annual Conference of the International Speech  
456 Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010*, pages 2794–2797, 2010b.
- 457 B. W. Schuller, S. Steidl, A. Batliner, A. Vinciarelli, K. R. Scherer, F. Ringeval, M. Chetouani, F. Weninger, F. Eyben,  
458 E. Marchi, M. Mortillaro, H. Salamin, A. Polychroniou, F. Valente, and S. Kim. The INTERSPEECH 2013  
459 computational paralinguistics challenge: social signals, conflict, emotion, autism. In *INTERSPEECH 2013, 14th  
460 Annual Conference of the International Speech Communication Association, Lyon, France, August 25-29, 2013*,  
461 pages 148–152, 2013.
- 462 C. Segura, D. Balcells, M. Umbert, J. Arias, and J. Luque. Automatic speech feature learning for continuous prediction  
463 of customer satisfaction in contact center phone calls. In A. Abad, A. Ortega, A. Teixeira, C. García Mateo, C. D.  
464 Martínez Hinarejos, F. Perdigão, F. Batista, and N. Mamede, editors, *Advances in Speech and Language Technologies  
465 for Iberian Languages*, pages 255–265, Cham, 2016. Springer International Publishing. ISBN 978-3-319-49169-1.
- 466 C. Vaudable and L. Devillers. Negative emotions detection as an indicator of dialogs quality in call centers. In *2012  
467 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5109–5112, 2012.  
468 doi:10.1109/ICASSP.2012.6289070.
- 469 O. Viikki and K. Laurila. Cepstral domain segmental feature vector normalization for noise robust speech recognition.  
470 *Speech Communication*, 25(1):133–147, 1998. ISSN 0167-6393. doi:[https://doi.org/10.1016/S0167-6393\(98\)00033-8](https://doi.org/10.1016/S0167-6393(98)00033-8).  
471 8.
- 472 G. Zweig, O. Siohan, G. Saon, B. Ramabhadran, D. Povey, L. Mangu, and B. Kingsbury. Automated quality monitoring  
473 for call centers using speech and nlp technologies. In *Proceedings of the 2006 Conference of the North American  
474 Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume:  
475 Demonstrations, NAACL-Demonstrations '06*, page 292–295, USA, 2006. Association for Computational Linguistics.  
476 doi:10.3115/1225785.1225796.

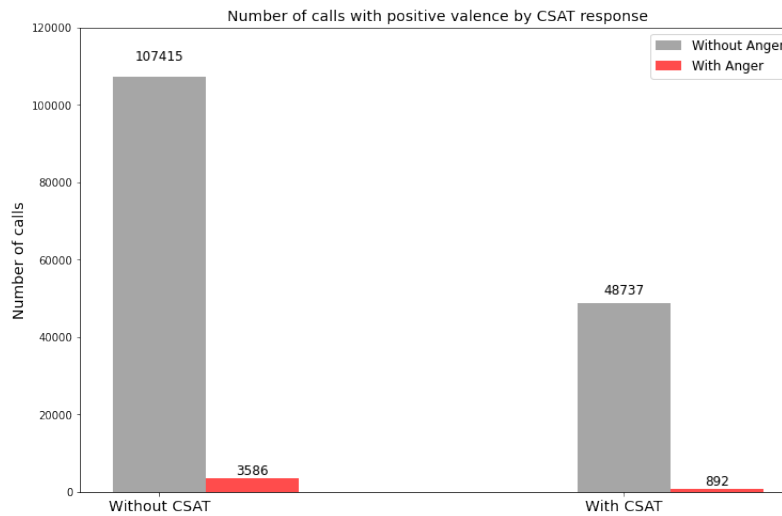




(a) CSAT scoring with respect to positive valence detection.



(b) CSAT scoring with respect to negative valence detection.



(c) CSAT scoring in respect to anger detection.

Figure 2: Number of calls with and without CSAT scoring in respect to detected emotions: (a) positive valence, (b) negative valence, (c) anger

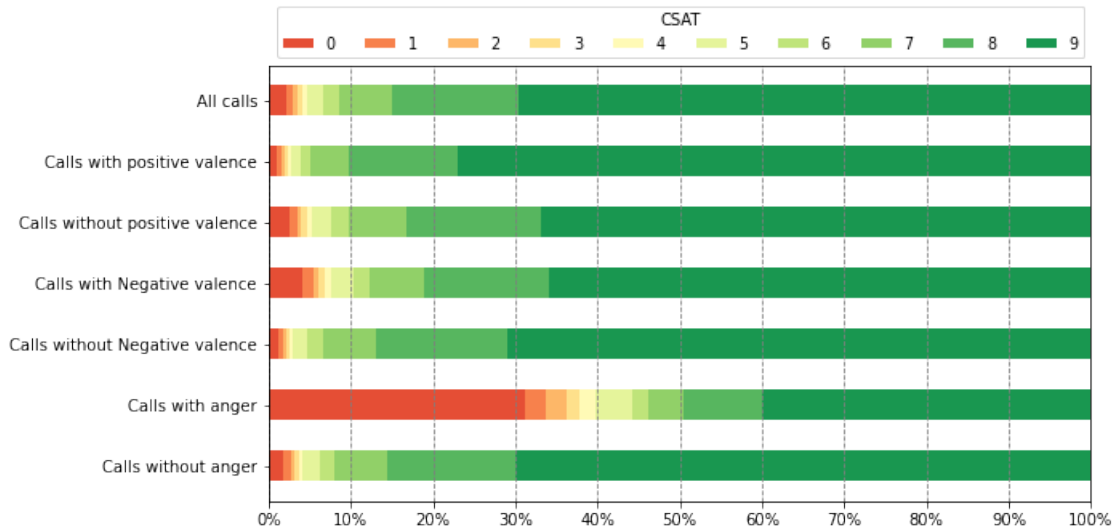


Figure 3: CSAT scoring distribution with respect to emotional analysis. Number of calls are reported in table 6

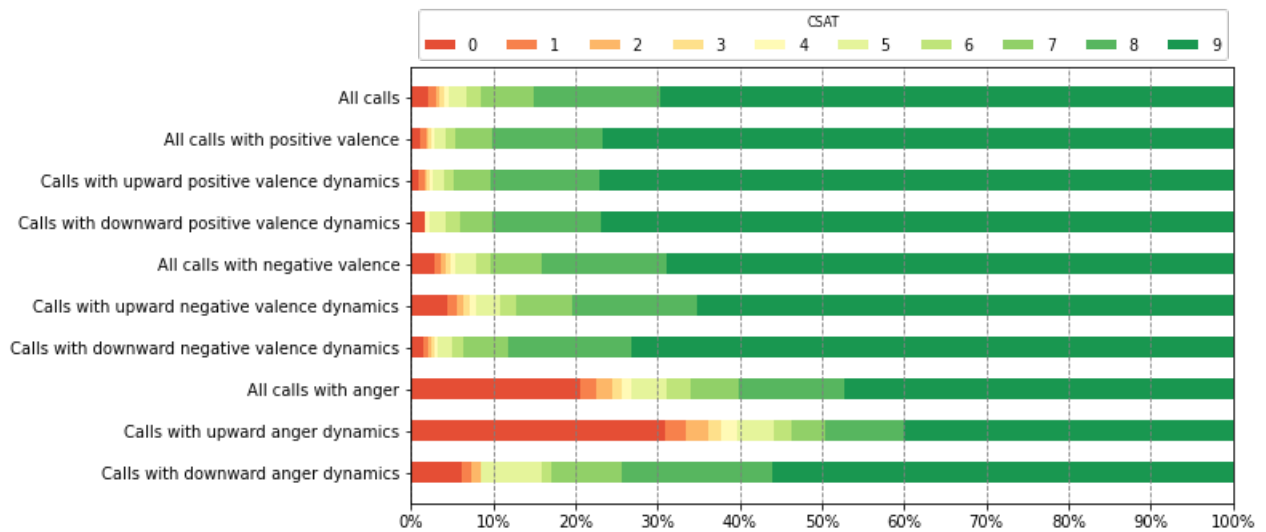


Figure 4: CSAT scoring distribution with of customer profiles.