Effect of gender and call duration on customer satisfaction in call center big data

Quim Llimona, Jordi Luque, Xavier Anguera, Zoraida Hidalgo, Souneil Park, Nuria Oliver

Telefonica Research
Edificio Telefonica-Diagonal 00, Barcelona, Spain
jls@tid.es

Abstract

Customer center call data is typically collected by organizations and corporations in order to improve customer experience through the analysis of such call data. In this paper, we report our findings when analysing more than 26 thousand calls to the call centers of a large corporation in a Latin American country. We focus on the impact of gender and call duration on self-reported customer satisfaction. Speech-based gender detection technology is employed to automatically detect the gender of the customer and the agent involved in the calls. A significant correlation is found between self-reported customer satisfaction at the end of the call and gender homophily between the customer and the call center's agent. Interestingly, we do not find any significant effect of call duration on satisfaction.

Index Terms: conversation, gender recognition, dialogue analysis interaction, speaker trait, computational paralinguistic

1. Introduction

In many countries and industries, customer service call centers are the main tool for customers to interact with a particular company. Moreover, the value of the data collected in call centers goes beyond traditional customer relationship management (CRM) as it also offers useful information, insights, and trends –both from existing and potential customers– that support sales and marketing. As such, call centers frequently implement advanced processes for service representative training, call recording and analysis, quality monitoring, and customer feedback acquisition. Recent advances in big data and speech analysis have opened up unprecedented opportunities for further innovation in this area. As companies grow and encounter diverse customer requirements, the creative application of such advanced technologies is actually becoming an imperative [1, 2, 3, 4].

One such area of innovation in call center data analytics is intelligent call routing based on modeling a variety of dimensions of both the customer and the agent [5, 6].

In particular, several studies have analysed the impact of demographic features –of both the agent and the customer– on customer service. In [7] demographic factors were taken into account to identify and predict the performance of the agent and Higgs *et al.* [8] found a significant relationship between age and performance in a sample of 289 call agents working in the UK. In their study, older call agents demonstrated better customer service skills than younger call agents, although there was no significant correlation between level of experience and job performance. Håkan *et al.* [9] reported on differences in verbal behavior, showing that female callers are more verbose than male callers, speaking more freely to describe their reason for calling. Moshavi and Terborg [10] found no performance differences between temporary and regular call agents.

Additional work has applied computational analysis to recorded call center conversations. In particular, scholars have explored the automatic prediction of call quality and customer satisfaction and have reported promising results. Park and Gates [11] have proposed machine learning techniques using linguistics and prosodic features to predict customer satisfaction. Zweig *et al.* [12] conducted a similar study using a number of manually selected features and a maximum entropy classification method. Recently, emotion in call center conversations has arisen as a research topic. Devillers *et al.* [13] have worked on detecting three emotional states (Anger, Positive, and Neutral) with acoustic features. They have studied data from two different call centers, service maintenance and medical emergency, and analysed the possibility to generalize a method trained on one of them to the other. Vaudable and Devillers [14] worked on detecting negative emotion in call center conversations and analysed its usefulness to predict customer satisfaction.

In this paper, we analyse more than 26 thousand phone conversations made to the call center of a major company in a Latin American country. All the calls were made in Spanish. The major contribution of our work is in the large scale paralinguistic analysis of the call center data. The analysis reveals novel findings about the relation between gender, call duration, and customer satisfaction. We are particularly interested in understanding whether there is an *homophily effect* on call center phone conversations. Homophily is the tendency for similar individuals to associate and is one of the most robust findings in social science. While homophily is related to a natural tendency of humans to link up to people similar to them, in this paper we focus on the impact of gender homophily in self-reported customer satisfaction in the context of a call center. In addition, our study presents an useful application scenario for automatic gender recognition techniques. In many call centers, the gender information of the customer database may not match with the actual caller's gender, for example, there are often multiple people living in the same household. We show the importance of using a speech-based gender detection [15, 16], instead of the gender information from the call center's database.

2. Automatic gender classification

Although the company running the call center may have demographic information about the customer, the person who makes the actual call to the call center may be different from the customer registered in their database. Therefore, we propose to use a voice-based automatic gender classification method.

A Gaussian SuperVector (GSV) based system similar to that in [16] is chosen due its competitive performance on gender detection tasks [15]. Two separate corpora are used to develop our method: a German spontaneous telephone speech database for training and an unseen Spanish dataset for testing. Performance on a small portion of the call center data is also reported.

2.1. Corpora

The training data is taken from the German SpeechDat II [17] corpus which is annotated with gender and age labels as given by callers at the time of recording. It mainly contains read speech including a set of words, sentences and digits. Recordings last between less than one second up to several seconds for phonetically rich sentences. The data used was an age-balanced subset of native German speakers. The training and test sets employed in [15] are combined to create a unique training set which leads to a total of 60 speakers per gender class and roughly 44 utterances per speaker. Children are removed from the data set. Youth and senior speakers are combined to create the two gender groups.

Two data sets are used to assess the performance and generalizability of the developed classifier. First, the Fisher Spanish - Speech corpus [18] –developed by the Linguistic Data Consortium– is used. It consists of 819 telephone conversations lasting from 10 to 12 minutes each, yielding roughly 163 hours of telephone speech from 136 native Caribbean Spanish and non-Caribbean Spanish speakers. A broad set of topics covered in the conversations ensure speech variability. Speaker segmentation is done by analysing independently each conversation channel, which is supposed to correspond to one speaker. It leads to a an unbalanced set of $1,638$ gender samples, $1,273$ female and 365 male. Second, we sample 50 calls from the call-center data set and manually annotate them with the gender information.

2.2. System Description

In our method, gender classification is performed in two steps. First, speech activity detection (SAD) is carried out by removing silent segments. Second, a binary classification task is performed using a Support Vector Machine (SVM) model trained on the Gaussian's means of a GMM model obtained from MAP-adapting a UBM model to the data.

2.2.1. Speech vs non-speech detection

First we compute a simple bi-Gaussian model of the log energy distribution for each conversation side. Speech activity detection (SAD) is performed to detect low-energy and highly likely non-speech frames and to remove them from further processing. The segmentation of speech *vs* non-speech frames is post-processed to eliminate noisy segments by removing (1) regions with simultaneous speech activity in both channels and (2) segments shorter than 3 seconds. The segmented speech frames are the input to the gender classifier.

2.2.2. Gender classification

We train gender-dependent Universal Background Models (UBMs) on German SpeechDat II telephone data. The two gender-dependent UBMs are incrementally trained up to 256 Gaussians. The GSV-UBM system is based on the gender and age recognition system proposed in [16]. Mel Frequency Cepstral Coefficients (MFCC) are employed to parametrize the speech signal. They are then modeled using diagonal covariance matrices and MAP adaptation technique to adapt the mean GMM-components from the UBM model. Next, we create a GMM supervector by stacking all 39-dimensional mean vectors. Supervectors y are built for every speaker by assigning them a gender label l. These supervectors are placed as support vectors and finally fed to a Support Vector Machine (SVM) classifier. In testing, a score or posterior class probability $Pr(y = l|x)$ is computed as the distance to the separation hyperplane, approximated by a sigmoid function [19]. Note that gender classification scores lower than 0.2 are discarded.

	60 s.	180 s.	300 s.	480 s.	full
w.o SAD	93.40 %	90.53 %	88.94 %	84.12 %	82.29 %
SAD	91.82 %	**94.38 %**	92.91 %	89.43 %	84.37 %

Table 1: Automatic gender classification on Fisher Spanish Speech corpus. Performance in terms of % weighted accuracy depending on call duration. Each column consists of $1,638$ *trials. Each row compares benefit on using silence removal (SAD).*

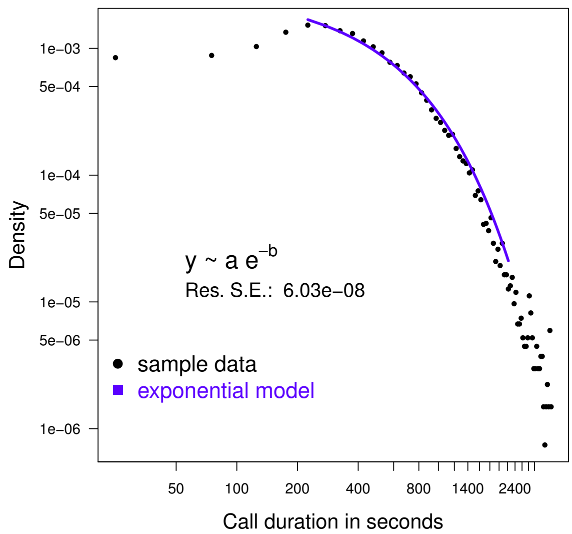

Figure 1: Density histogram in log-log of phone call durations for $26,881$ *calls and corresponding exponential fitting. Fitting is performed on the plotted blue line range. Mean call duration is around* 500 *seconds.*

2.2.3. Empirical evaluation

Table 1 reports the accuracy of our gender classifier on the Fisher corpus phone data for different chunk durations. Audio channels are trimmed in chunks of $1, 3, 5, 8$ minutes in order to assess how trial duration affects accuracy. As can be noticed, the GSV-GMM system suffers from the mismatch of duration between training and testing data, an effect previously reported on the literature [20]. We report accuracies with and without SAD. Given the results on Table 1, we select the model with SAD and 3-minute chunk duration (94.38% accuracy). In the remainder of this paper, we report results of this model and always use inferred gender as opposed to the gender that appears in the call center's database, for reasons explained below.

In order to validate generalization and assess the performance on the call center data, 4 different people annotated gender in 50 conversations of the dataset under study, with an inter-rater agreement index of Fleiss' Kappa $= .97$. Our proposed gender classifier obtained 99% accuracy when compared

	Database gender	
Inferred gender	Female	Male
Female	2,334	910
Male	656	4,820
Sum	2,890	5,730

Table 2: Confusion matrix between gender as it appears on the call center's database (columns) and inferred gender as per our algorithm (rows). Only customer's gender.

	Gender (agent/customer)			
	F/F	F/M	M/F	M/M
counts	3,781	7,583	2,451	5,047

Table 3: Number of phonecalls per gender duple involved in the conversation (Low confidence duples, $7,080$ calls are removed)

to only 81% accuracy of the gender as it appears in the call center's database. Table 2 reports the confusion matrix in both cases showing a mismatch of around 20% between both information sources. Note that no call center information about customer gender is provided in a large number of calls (around 65%). Given the reported accuracy of our algorithm of 94.38%, this large error is most likely due to "calls on behalf of" the customer that appears in the call center's database. For this reason, in the following we use the gender inferred by our algorithm instead of the gender that appears in the call center's database. In the following, detected gender is assigned to all calls remaining next to confidence and error removal, see Table 3.

3. Effect of gender and call duration on satisfaction

The goal of our research is to shed light on the effect of call duration and gender homophily on customer satisfaction. Our aim is to identify variables that could be relevant in the design of a call center's operations in order to increase customer satisfaction. We first describe the call center data that is used in our study, followed by an analysis of the impact of call duration and gender homophily on self-reported satisfaction.

3.1. Call center data

The call center data is composed of $26,881$ inbound phone calls. It represents a random subset of calls extracted from contact centers in one Latin American country. All the calls were made in Spanish. Data was collected throughout one month such that it comprises a variety of interactions between the customer and the call center's agent. First, we run our gender classification algorithm to automatically segment the gender of the customer and the agent, generating a gender duple $(gender_{agent}, gender_{customer})$ per conversation. Table 3 reports the amount of calls for each possible gender duple.

Figure 1 depicts a log-log density histogram of phone call duration and an exponential fit of the data. It has a mean and a median value of around 500 and 400 seconds, respectively. Note that we compared several power-law and log-normal fittings but obtained worst results in terms of residual square error and Kolmogorov-Smirnov statistic [21]. This results contrasts previous work on modeling call duration distributions for individual users on mobile networks [22, 23], where TLAC and log-normal distributions have been found to best fit the data in larger datasets. It is worth to note that duration call values employed in our analysis were those provided by the call center.

At the end of each call, the customer is called back and gently asked to complete a answer a question related to the service: *According to its previous call to our call center, how satisfied, overall, are you with the telephone service of XXXX. Press 1-5 where 1 is very dissatisfied and 5 very satisfied.*

	Satisfaction					
	0	1	2	3	4	5
counts	9572	2420	580	1113	2155	11041

Table 4: Number of calls with respect to self-reported satisfaction.

Table 4 reports the distribution of customer satisfaction in our dataset. Note that not every customer replies to the satisfaction survey and that repeated calls were removed from the population sampling. In such cases, a satisfaction value of 0 is assigned and removed from further analysis. For that reason, the total number of calls containing information related to satisfaction decreased to $17,309$ calls from the initial set. The distribution is significantly skewed towards 5 (64%). Calls with low level of satisfaction, 1 and 2, are approximately 10% of the data.

3.2. Gender, call duration and satisfaction

A Pearson's Chi Squared test of independence [24] is employed to investigate the relationship between gender, call duration and the customer's satisfaction with the call. When the null hypothesis is true then the probability distribution of a given statistic, based on a random sample, follows a chi-square distribution. The $\tilde{\chi}^2$ statistic provides an index of the "fit" between the hypothesized distribution and the real population relative frequency distribution:

$$\tilde{\chi}^2 = \sum_{k=1}^{N} \frac{(O_k - E_k)^2}{E_k} \tag{1}$$

where O_k stands for observed frequencies in each category k and E_k for expected frequencies, that is, the number of observations in the sample that should be in the category if the null hypothesis were correct. When the null hypothesis is correct, there is no statistically significant difference between the observed and the expected frequencies: $\tilde{\chi}^2$ is in the rejection region and then the p-value is higher than the α (the level of acceptable error rate). In all tests simulated p-values are estimated on 10^5 replicates with Monte Carlo simulation [1].

Table 5 reports the observed and the expected distributions (in parenthesis) in terms of satisfaction and gender homophily between the customer and the agent. The last row contains the sum of the total number of samples in each self-reported satisfaction class. Satisfaction levels 1 and 2 are combined to create a "Low" satisfaction category. The same procedure is applied to scores 4 and 5 which then are mapped to the "High" satisfaction category. A score of 3 is kept as "Neutral". Results from Table 5 support the rejection of the null hypothesis, that is, the level of satisfaction differs for the different groups depending on the agent-customer gender duple ($\tilde{\chi}^2(N = 14,316) = 14.07, p = .023$). In particular, when the

[1] The package 'stats' from R[25] statistical software has been used.

	Satisfaction		
Match (agent/customer)	Low	Neutral	High
Female-Female	465 (508.3)	188 (182.6)	2,258 (2,220)
Male-Male	715 (668.8)	242 (240.2)	2,873 (2,920.9)
Female-Male	1,019 (992.6)	338 (356.6)	4,327 (4,334.9)
Male-Female	301 (330.2)	130 (118.6)	1,460 (1,442.1)
Sum	2,500	898	10,918

.

Table 5: Observed and expected distributions of satisfaction depending on the agent/customer inferred genders (in parenthesis). The level of satisfaction differs for the different groups depending on gender $\tilde{\chi}^2(N = 14,316) = 14.07, p = .023$

	Call duration	
Match (agent/customer)	Short	Long
Male-Male	1,792 (1,771.8)	3,124 (3,144.2)
Female-Female	1,339 (1,324.2)	2,335 (2,349.8)
Female-Male	2634 (2,664.2)	4,758 (4,727.8)
Male-Female	849 (853.8)	1,520 (1,515.2)
Sum	6,614	11,737

Table 6: Contingency table for observed and expected distributions (in parenthesis) of call duration depending on the agent/customer inferred genders. The null hypothesis is not rejected, that is, the duration of phone calls does not differ between the different gender combinations involved in the phone call $\tilde{\chi}^2(N = 18,351) = 1.20, p = .748$

gender of the agent matches the customer's agent, there would be higher probability than expected for the customer to be satisfied with the call. This finding supports the concept of the *homophily effect* [26] which argues that people have a preference to interact with people who are similar to them. According to results depicted on Table 7, the level of satisfaction also significantly differs depending on the customer's gender ($\tilde{\chi}^2(N = 13,418) = 10.65, p < .01$), but to a lesser degree than gender homophily.

In order to investigate the relationship between call duration and gender, we first classify the calls into *short* (lasting between $[50, 300]$ seconds) and *long* (lasting > 300 seconds) calls. Selection of *short* calls is based on upper threshold fixed to mean duration value minus half of a standard deviation (around 300 seconds). Given that we discard phone calls that are shorter than 50 seconds ($6,614$ phone calls), there is a total of $18,351$ phone calls left for the analysis. Table 6 contains the observed and the expected distributions from the $\tilde{\chi}^2$ statistic. In this case, the null hypothesis is not rejected as there is no significant difference in phone call duration depending on the different genders involved in the dialogue. ($\tilde{\chi}^2(N = 18,351) = 1.20, p = .748$). Taking into account only the customer's gender and duration of the call – Table not reported –, as in the previous scenario, the null hypothesis is not rejected suggesting that customer gender does not affects call duration $\tilde{\chi}^2(N = 18,351) = 0.107, p = .739$. Finally Table 8 contains the contingency table for self-reported customer satisfaction and call duration. In this case the null hypothesis is not rejected either $\tilde{\chi}^2(N = 13,924) = 1.599, p = .452$, that is, there is no evidence in the analysed data that call duration impacts self-reported satisfaction. However, we do not make a conclusion yet since call duration is likely to be associated more strongly with other factors such as the reason of the call, complexity of the problem, previous experience of the customer, etc. We plan to investigate the effects of such factors together as our next step.

	Satisfaction	
Customer's gender	Low	High
Male	1,734 (1,664.6)	7,200 (7,269.5)
Female	766 (835.5)	3,718 (3,648.6)
Sum	2,500	10,918

Table 7: Observed and expected distributions (in parenthesis) depending on the customer's inferred gender. The duration of the phone calls differs for the different groups $\tilde{\chi}^2(N = 13,418) = 10.65, p < .01$ *(Yates' correction)*

	Call Duration	
Satisfaction	Short	Long
Low	880 (874.72)	1,539 (1544.3)
Neutral	298 (315.3)	574 (556.7)
High	3,857 (3844.9)	6,776 (6788.0)
Sum	5,035	8,889

Table 8: Observed and expected distributions (in parenthesis) of satisfaction with respect to call duration. The null hypothesis is not rejected, that is, the level of satisfaction does not differ with call duration $\tilde{\chi}^2(N = 13,924) = 1.599, p = .452$

4. Conclusions

In this paper, we have explored the relationship between three relevant variables in customer care within a call center scenario: gender (of both the customer and the agent), call duration and customer satisfaction with the phone call.

As a first step, we have developed an automatic voice-based gender identification method which is able to correctly classify gender with high accuracy on several data sets. Interestingly, we find a significant difference between the customer's gender as it is recorded in the call center's database and the gender of the calling customer. Note that typically there are multiple individuals living in a household and it is not guaranteed that the calling customer would be the same as the customer registered in the call center's database. Therefore, in the remainder of our analysis we have used the gender as it is inferred by our proposed algorithm. We have found empirical evidence that gender homophily plays a significant role on self-reported customer satisfaction in a call center. It is worth to note that we have just found a correlation and not a causal relationship. Interestingly, we have not identified any significant effects of call duration on self-reported levels of satisfaction.

Given the importance of gender homophily on customer satisfaction and the fact that in about 20% of call center calls the calling customer's gender does not match the call center's database information, call centers would benefit from on-line algorithms to automatically infer the customer's gender.

5. Acknowledgements

This project has received funding from the EU's Horizon 2020 research and innovation programme under grant agreement No 645323. This text reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

6. References

[1] L. Bocklund and D. Bengtson, *Call Center Technology Demystified: The No-nonsense Guide to Bridging Customer Contact Technology, Operations and Strategy*. Call Center Press, 2002. [Online]. Available: http://books.google.ca/books?id=jtAHAAAACAAJ 1

[2] S. Kopparapu, *Non-Linguistic Analysis of Call Center Conversations*, ser. SpringerBriefs in Electrical and Computer Engineering. Springer International Publishing, 2014. [Online]. Available: http://books.google.es/books?id=IecwnwEACAAJ 1

[3] G. Mishne, D. Carmel, R. Hoory, A. Roytman, and A. Soffer, "Automatic analysis of call-center conversations," in *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, ser. CIKM '05. New York, NY, USA: ACM, 2005, pp. 453–459. [Online]. Available: http://doi.acm.org/10.1145/1099554.1099684 1

[4] S. Busemann, S. Schmeier, and R. G. Arens, "Message classification in the call center," in *Proceedings of the Sixth Conference on Applied Natural Language Processing*, ser. ANLC '00. Stroudsburg, PA, USA: Association for Computational Linguistics, 2000, pp. 158–165. [Online]. Available: http://dx.doi.org/10.3115/974147.974169 1

[5] A. R. Ali, "Intelligent call routing: Optimizing contact center throughput," in *Proceedings of the Eleventh International Workshop on Multimedia Data Mining*, ser. MDMKDD '11. New York, NY, USA: ACM, 2011, pp. 5:1–5:9. [Online]. Available: http://doi.acm.org/10.1145/2237827.2237832 1

[6] Z. Aksin, M. Armony, and V. Mehrotra, "The modern call center: A multi-disciplinary perspective on operations management research," *Production and Operations Management*, vol. 16, no. 6, p. 665688, 2007. 1

[7] D. Kwok and U. of Western Australia. Graduate School of Education, *Profiles of High-performing Female Call Centre Agents*. University of Western Australia, 2007. [Online]. Available: http://books.google.es/books?id=m_FPPgAACAAJ 1

[8] N. Shamsuddin and R. A. Rahman, "The relationship between emotional intelligence and job performance of call centre agents," *Procedia - Social and Behavioral Sciences*, vol. 129, no. 0, pp. 75 – 81, 2014, 2nd International Conference on Innovation, Management and Technology Research. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042814028328 1

[9] H. Jonsson and R. Eklund, "Gender differences in verbal behaviour in a call routing speech application," in *Proceedings from Fonetik 2011, Quarterly Progress and Status Report TMH-QPSR, Volume 51, 2011*. Universitetsservice, 2011, pp. 81–84. 1

[10] D. Moshavi and J. R. Terborg, "The job satisfaction and performance of contingent and regular customer service representatives," *International Journal of Service Industry Management*, vol. 13, no. 4, pp. 333–347, 2002. [Online]. Available: http://dx.doi.org/10.1108/09564230210445069 1

[11] Y. Park and S. C. Gates, "Towards real-time measurement of customer satisfaction using automatically generated call transcripts," in *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009, pp. 1387–1396. 1

[12] G. Zweig, O. Siohan, G. Saon, B. Ramabhadran, D. Povey, L. Mangu, and B. Kingsbury, "Automated quality monitoring for call centers using speech and nlp technologies," in *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume: demonstrations*. Association for Computational Linguistics, 2006, pp. 292–295. 1

[13] L. Devillers, C. Vaudable, and C. Chastagnol, "Real-life emotion-related states detection in call centers: a cross-corpora study," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010. 1

[14] C. Vaudable and L. Devillers, "Negative emotions detection as an indicator of dialogs quality in call centers," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5109–5112. 1

[15] F. Metze, J. Ajmera, R. Englert, U. Bub, F. Burkhardt, J. Stegmann, C. Muller, R. Huber, B. Andrassy, J. Bauer, and B. Littel, "Comparison of four approaches to age and gender recognition for telephone applications," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4, April 2007, pp. IV–1089–IV–1092. 1, 2

[16] T. Bocklet, A. Maier, J. Bauer, F. Burkhardt, and E. Noth, "Age and gender recognition for telephone applications based on gmm supervectors and support vector machines," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, March 2008, pp. 1605–1608. 1, 2

[17] H. Hge, C. Draxler, H. van den Heuvel, F. T. Johansen, E. Sanders, and H. S. Tropf, "Speechdat multilingual speech databases for teleservices: Across the finish line," in *PROC. EUROP. CONF. SPEECH PROC. AND TECHN. (EUROSPEECH*, 1999, pp. 2699–2702. 2

[18] Graff, David, and et al., "Fisher spanish speech," *LDC2010S01. DVD. Philadelphia: Linguistic Data Consortium*, 2010. 2

[19] J. C. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," in *ADVANCES IN LARGE MARGIN CLASSIFIERS*. MIT Press, 1999, pp. 61–74. 2

[20] C. HanilçI and F. Ertaş, "Investigation of the effect of data duration and speaker gender on text-independent speaker recognition," *Comput. Electr. Eng.*, vol. 39, no. 2, pp. 441–452, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.compeleceng.2012.09.014 2

[21] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," *SIAM Review*, vol. 51, no. 4, pp. 661–703, 2009. [Online]. Available: http://dx.doi.org/10.1137/070710111 3

[22] P. O. S. V. De Melo, L. Akoglu, C. Faloutsos, and A. A. F. Loureiro, "Surprising patterns for the call duration distribution of mobile phone users," in *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III*, ser. ECML PKDD'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 354–369. [Online]. Available: http://dl.acm.org/citation.cfm?id=1889788.1889812 3

[23] L. Brown, N. Gans, A. Mandelbaum, A. Sakov, H. Shen, S. Zeltyn, and L. Zhao, "Statistical analysis of a telephone call center," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 36–50, 2005. [Online]. Available: http://dx.doi.org/10.1198/016214504000001808 3

[24] A. Glenberg and M. Andrzejewski, *Learning from Data: An Introduction to Statistical Reasoning*. Routledge, 2014. [Online]. Available: http://books.google.es/books?id=09eyngEACAAJ 3

[25] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/ 3

[26] P. F. Lazarsfeld and R. K. Merton, "Friendship as a social process: A substantive and methodological analysis," in *Freedom and Control in Modern Society*, M. Berger, T. Abel, and C. Page, Eds. New York: Van Nostrand, 1954, pp. 18–66. 4