Uma abordagem de aprendizagem semissupervisionada para a classificação automática de personalidade baseada em pistas acústico-prosódicas

Rubén Solera-Ureña[1], Helena Moniz[1,2,3], Fernando Batista[1,4], Vera Cabarrão[1,2], Anna Pompili[1,5], Ramón Fernández-Astudillo[1,6], Isabel Trancoso[1,5]

[1] Laboratório de Sistemas de Língua Falada, INESC-ID Lisboa, Lisboa, Portugal
[2] FLUL/CLUL, Universidade de Lisboa, Lisboa, Portugal
[3] Unbabel Lda, Lisboa, Portugal
[4] Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
[5] Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal
[6] IBM Research AI, Yorktown Heights, NY, USA

Abstract:
Automatic personality analysis has gained great attention in the last years as a fundamental dimension in human-machine interactions. However, the development of this technology in some domains, such as the classification of children's personality, has been hindered by the limited number and size of the available speech *corpora* due to ethical concerns on collecting such *corpora*. To circumvent the lack of data, we have investigated the application of a semi-supervised training approach that makes use of heterogeneous (age and language mismatches) and partially non-labelled data sets. Namely, preliminary personality models trained using a small labelled data set with French speaking adults are iteratively refined using a larger unlabeled set of Portuguese children's speech, whereas a labelled *corpus* of Portuguese children is used for evaluation. We also investigated speech representations based on prior linguistic knowledge on acoustic-prosodic clues for personality classification tasks and have analysed their relevance in the assessment of each personality trait. The results point out to the potential of applying semi-supervised learning approaches with heterogeneous data sets to overcome the lack of labelled data in under-resourced domains, and to the existence of acoustic-prosodic clues shared by speakers with different languages and ages, which allows for the classification of personality independently of these variables.

Keywords: *computational paralinguistics, automatic personality classification, cross-language, cross-age, acoustic-prosodic features*
Palavras-chave: análise paralinguística computacional, classificação automática de personalidade, línguas distintas, faixas etárias diferentes, pistas acústico-prosódicas

1. Introdução

As intenções, emoções e até os traços de personalidade são passíveis de serem codificados na fala como informação paralinguística, ou seja, como estando para além das estruturas linguísticas da língua. Este tópico tem merecido a atenção de investigadores em distintas áreas nos últimos anos. A análise dos traços de personalidade, em particular, tem aplicações diversas tanto em comunicações pessoa-pessoa como em interações pessoa-máquina, como a identificação de traços de personalidade para a discriminação de comportamentos sociais em crianças autistas a interagirem com outras crianças ou com agentes virtuais/robôs.

O presente trabalho foi financiado pela Fundação para a Ciência e a Tecnologia (FCT), referência UID/CEC/50021/2019, pelos projetos INSIDE (referência CMUP-ERI/HCI/0051/2013) e BioVisualSpeech (referência CMUP-ERI/TIC/0033/2014), e pelo financiamento das bolsas de doutoramento SFRH/BD/96492/2013 e SFRH/BD/97187/2013 e de pós-doutoramento SFRH/PBD/95849/2013.

Grande parte da literatura atual sobre análise automática de personalidade concentra-se principalmente na avaliação ou deteção desses traços, com base em vários conjuntos diferentes de pistas na fala. No entanto, as aplicações de inteligência artificial têm vindo a utilizar, na última década, métodos de aprendizagem automática para, de forma progressiva, dotar os robôs e agentes virtuais da capacidade de simular alguns traços emocionais e detetar automaticamente as emoções e os traços de personalidade dos seus parceiros humanos. O objetivo é promover uma interação eficaz e tornar a comunicação mais idiossincrática e sintonizada com as particularidades paralinguísticas dos seus interlocutores. A tipologia baseada nos traços de personalidade denominados ***Big-Five*** (**OCEAN**) corresponde a um modelo psicológico amplamente utilizado, que visa descrever a personalidade humana em termos de cinco grandes dimensões: ***Openness* ou abertura** (pessoa curiosa, imaginativa, perspicaz, original, com interesses amplos), ***Conscientiousness* ou conscienciosidade** (eficiente, organizada, confiável, responsável), ***Extroversion* ou extroversão** (ativa, assertiva, energética, extrovertida, faladora), ***Agreeableness* ou amabilidade** (gentil, generosa, simpática, confiante) e ***Neuroticism* ou neuroticismo** (ansiosa, tensa, sensível, instável, preocupada).

A classificação automática da personalidade é ainda uma tarefa muito desafiante, quer pelo espectro individual do falante, quer pelo espectro dos próprios traços de personalidade: ainda que seja comummente assumido que a complexidade de uma pessoa possa ser definida pelo modelo *Big-Five* através de cinco dimensões de personalidade, tal tipologia pode não cobrir todas as sub-especificações ou as fronteiras entre tais dimensões, como alguns estudos psicológicos já têm vindo a apontar (e.g., Cloninger, 2009). Também é sabido que alguns traços podem ser mais facilmente reconhecidos através de procedimentos automáticos do que outros (por exemplo, reconhecimento de padrões e métodos de aprendizagem automática), mas isso pode variar de acordo com os dados de fala disponíveis e as metodologias empregues (veja-se Vinciarelli & Mohammadi (2014), para uma revisão sobre abordagens computacionais para a deteção de personalidade). Além disso, tem sido apontado que diferentes traços de personalidade são revelados na fala espontânea através de diferentes conjuntos de pistas acústico-prosódicas (Vinciarelli & Mohammadi, 2014; Mairesse et al., 2007; Polzehl, Moeller & Metze, 2010a; Polzehl, Moeller & Metze, 2010b; Polzehl, Moeller & Metze, 2011), mas ainda são muito escassas as categorizações exaustivas de tais características e os estudos sobre o seu impacto em termos de variáveis como idade, língua, cultura, etc.

Importa também acrescentar que a classificação computacional de personalidade é um campo de investigação recente, sendo que os conjuntos de dados de fala anotados em termos de personalidade são escassos e assaz pequenos (Vinciarelli & Mohammadi, 2014). Este facto é ainda mais acentuado para determinadas línguas com relativamente poucos recursos e em áreas específicas, como é o caso da classificação da personalidade de crianças portuguesas. Essa falta de dados é um dos obstáculos mais importantes na análise de paralinguística computacional, uma vez que dificulta seriamente o desenvolvimento de modelos de personalidade robustos e precisos por meio de métodos de aprendizagem automática. Infelizmente, o desenvolvimento de novos *corpora* de fala para tarefas de classificação de personalidade, devidamente anotados e com dimensão suficiente para permitir o treino de modelos robustos, é uma tarefa dispendiosa, morosa e que levanta problemas de privacidade e proteção das crianças, sendo mesmo proibida em muitos casos por questões éticas (Vinciarelli & Mohammadi, 2014). Pelas razões apresentadas, são necessárias novas abordagens para contornar a escassez de dados de treino nessas circunstâncias e assim desenvolver aplicações para a classificação computacional de personalidade em cenários de inteligência artificial.

Os trabalhos anteriores nesta área geralmente estabelecem uma configuração experimental com apenas um único conjunto anotado de dados de fala, com características similares em termos de idade e língua dos falantes (Mohammadi & Vinciarelli, 2012; Schuller et al., 2012; Vinciarelli & Mohammadi, 2014; Schuller et al., 2015). Como referido anteriormente, o recurso a apenas um conjunto de dados poderá restringir a precisão e robustez dos modelos e, portanto, limitar o seu desempenho e uso em aplicações reais. Numa perspetiva

divergente dos estudos anteriores, a abordagem apresentada neste trabalho fundamenta-se numa configuração experimental que utiliza diversos *corpora*. Pretende-se, assim, investigar o uso de vários conjuntos de dados de fala para treinar modelos robustos, mesmo que estes tenham características marcadamente diferentes (por exemplo, disparidade entre idade e língua dos falantes) ou mesmo que não tenham sido anotados com traços de personalidade. Os fundamentos teóricos e metodológicos para esta abordagem são os que a seguir se descrevem. Primeiramente, estudos psicológicos longitudinais têm vindo a debater a dicotomia entre mudança *vs.* continuidade nos traços de personalidade desde a infância até à idade adulta (Kagan, 1996; Lewis, 1997; Roberts, Walton & Viechtbauer, 2006). Estudos como Caspi et al. (2003) mostram mesmo que os traços de personalidade das crianças estão intimamente ligados àqueles exibidos na idade adulta. Em segundo lugar, a forma como os falantes exibem as emoções e os traços de personalidade na fala tem uma natureza essencialmente paralinguística, detetável com recurso a pistas acústico-prosódicas (Mohammadi & Vinciarelli, 2012; Schuller et al., 2012; Vinciarelli & Mohammadi, 2014; Schuller et al., 2015). Com base nesses resultados, colocámos a hipótese da existência de conjuntos semelhantes de características acústico-prosódicas independentes da língua na fala de crianças e de adultos, que podem ser usadas para treinar modelos robustos para a classificação de traços de personalidade. Finalmente, o surgimento de metodologias semissupervisionadas e de transferência de aprendizagem proporcionam aos investigadores na área de aprendizagem automática procedimentos para lidar com conjuntos de dados parcialmente anotados (Chapelle, Schölkopf & Zien, 2006; Coutinho, Deng & Schuller, 2014; Deng, Zhang & Schuller, 2014). Esses procedimentos visam extrair conhecimento de informação explícita presente em pequenos conjuntos de dados anotados (aprendizagem supervisionada) e explorá-lo para extrair o conhecimento implícito em *corpora* não anotados de dimensões mais expressivas (aprendizagem não supervisionada). Tal permite desenvolver modelos com um desempenho superior àqueles que poderiam ser obtidos se treinados apenas com um pequeno conjunto de dados anotados.

Com base nas novas metodologias acima apresentadas, temos vindo a investigar a aplicação de uma abordagem de treino semissupervisionado (ou auto-aprendizagem) que faz uso de conjuntos de dados heterogéneos (disparidade de idade e língua dos falantes) e parcialmente anotados, com o objetivo de desenvolver modelos para a classificação automática de personalidade de crianças portuguesas. Uma outra linha de investigação compreende o uso de pistas acústico-prosódicas baseadas em conhecimento linguístico prévio, combinadas com conjuntos de pistas usados habitualmente em desafios internacionais de classificação de eventos paralinguísticos. Os nossos resultados anteriores apontam para taxas de desempenho razoáveis na classificação dos traços de personalidade abertura, extroversão e amabilidade (Solera-Ureña et al., 2016; Solera-Ureña et al., 2017).

Neste artigo, aprofundamos a análise dos resultados apresentados em trabalhos anteriores. Além disso, apresentamos um estudo detalhado sobre a relevância de pistas acústico-prosódicas específicas para a classificação de cada traço de personalidade no modelo *Big-Five*. O artigo está organizado da seguinte forma: a Secção 2 descreve as bases de dados de fala, as pistas acústico-prosódicas e os modelos de aprendizagem automática usados para a classificação automática de personalidade. Os resultados experimentais são apresentados na Secção 3. Finalmente, a Secção 4 apresenta as conclusões e direções de trabalho futuro.

2. Metodologia

2.1. *Corpora* de fala

Para a classificação automática de personalidade de crianças, foram utilizados distintos *corpora*. O *Speaker Personality Corpus* (SPC) (Mohammadi & Vinciarelli, 2012; Schuller et al., 2012; Schuller et al., 2015) foi usado neste trabalho para o treino de modelos estatísticos (classificadores binários) para cada traço

de personalidade do modelo *Big-Five*. O CNG *Corpus of European Portuguese Children's Speech* (CNG) (Hämäläinen et al., 2013), não anotado, foi usado a seguir numa abordagem semissupervisionada para refinar, de forma iterativa, os modelos iniciais. Finalmente, utilizou-se o *corpus Game-of-Nines* (GoN) (Campos, Alves-Oliveira & Paiva, 2016) como conjunto de teste, para estudar a transposição de modelos de personalidade construídos a partir de fala mista (adultos francófonos e crianças portuguesas) em modelos de traços de personalidade de crianças portuguesas. Nas subsecções seguintes, apresenta-se uma breve descrição desses três conjuntos de dados de fala.

2.1.1. *Speaker Personality Corpus*

A base de dados *Speaker Personality Corpus* (Mohammadi & Vinciarelli, 2012) consiste em 640 ficheiros de fala de 322 indivíduos suíços francófonos, recolhidos no boletim de notícias em francês da Radio Suisse Romande. Cada ficheiro contém 10 segundos de fala de um único falante (cerca de 1 hora e 40 minutos no total). Todos os ficheiros foram avaliados de forma independente por 11 anotadores (não falantes de francês), com recurso ao questionário de personalidade BFI-10 (Rammstedt & John, 2007). Para cada ficheiro, as etiquetas finais para os cinco traços de personalidade OCEAN são atribuídas por um procedimento de votação maioritária: para cada traço, é atribuído um nível alto ou baixo (denotado como O/NO, C/NC, E/NE, A/NA, N/NN, respetivamente), atendendo a que pelo menos 6 juízes pontuaram o ficheiro com os níveis alto/baixo.

O *corpus* SPC foi anteriormente usado no *Interspeech 2012 Speaker Trait Challenge-Personality Sub-challenge* (Schuller et al., 2012; Schuller et al., 2015). O *corpus* foi dividido em três subconjuntos de treino, desenvolvimento e teste, independentes do falante, com 256, 183 e 201 ficheiros, respetivamente. Neste trabalho, foi adotada a mesma configuração experimental de modo a estabelecer possíveis comparações com os resultados obtidos. A Tabela 1 ilustra o número de exemplos (ficheiros) em cada classe (nível alto ou baixo para um traço) do *corpus* SPC:

	SPC Treino					SPC Desenvolvimento					SPC Teste				
Traço	O	C	E	A	N	O	C	E	A	N	O	C	E	A	N
#Altos	97	110	121	139	140	70	81	92	79	88	80	99	107	105	90
#Baixos	159	146	135	117	116	113	102	91	104	95	121	102	94	96	111

Tabela 1: Número de exemplos em cada classe (nível alto/baixo para um traço) do *corpus* SPC.

2.1.2. CNG *Corpus of European Portuguese Children's Speech*

O CNG *Corpus of European Portuguese Children's Speech* (Hämäläinen et al., 2013) é um *corpus* oral de leitura e repetição desenhado originalmente para tarefas de reconhecimento automático de fala. Este *corpus* compreende cerca de 20 horas de fala de 484 falantes, organizados em dois subconjuntos diferentes com crianças de 3 a 6 anos e crianças de 7 a 10 anos, respetivamente. Dependendo da idade e das competências de leitura, as crianças leem em voz alta ou repetem exemplos diversos de quatro tipos diferentes de estruturas: frases foneticamente ricas (i.e., conjunto de frases que cobre todos os fonemas possíveis com uma distribuição aproximadamente uniforme), notas musicais ("dó", "ré", "mi"...), números cardinais isolados e sequências de números cardinais. Neste estudo, usamos apenas o subconjunto de frases foneticamente ricas proferidas pelas crianças de 7 a 10 anos (303 falantes, 6060 frases, 5,1 horas de fala). Esta escolha deveu-se ao facto de este ser o subconjunto mais semelhante aos dados alvo do estudo (conjunto de teste com crianças portuguesas com

idades entre os 10 e os 12 anos), bem como ao facto de as frases foneticamente ricas serem mais propensas a exibir traços de personalidade do que os outros tipos de frases do *corpus*.

2.1.3. *Game-of-Nines Corpus*

O *corpus Game-of-Nines* (Campos, Alves-Oliveira & Paiva, 2016) foi originalmente concebido para estudar a forma como o conflito se manifesta em interações sociais entre crianças, através da observação de pistas comportamentais (por exemplo, a fixação do olhar) numa interação com objetivos mistos (nomeadamente, um cenário com incentivos para a competitividade e a cooperação). Este inclui gravações sincronizadas de vídeo e áudio de 11 sessões diádicas com 22 crianças portuguesas (13 meninas e 9 meninos), com idades entre os 10 e os 12 anos, a jogar um jogo de cartas cooperativo (uma versão modificada do *Game of Nines* de Kelley, Linden Beckman & Fischer (1967)). A duração das gravações varia entre 9 a 18,6 minutos, com uma duração média de 12,8 minutos e um total de 2 horas e 20 minutos.

A base de dados original GoN foi pré-processada e adaptada aos nossos propósitos (Solera-Ureña et al., 2016). Em primeiro lugar, foi usada apenas a informação de áudio. Em segundo lugar, as transcrições das gravações foram usadas para identificar e extrair todos os segmentos de fala correspondentes a cada criança. O áudio correspondente à fala sobreposta foi removido. Como resultado desse pré-processamento, foram gerados três subconjuntos de fala diferentes:

- *GoN-complete*: todos os segmentos de fala de uma determinada criança durante a sessão foram concatenados num único ficheiro de fala. Como resultado, o subconjunto *GoN-complete* contém 22 ficheiros, com dimensões que variam dos 49 segundos a 8,1 minutos de fala (duração média de 4,2 minutos).
- *GoN-20seconds*: para cada criança, 4 ficheiros diferentes com cerca de 20 segundos de fala cada um foram gerados pela concatenação dos seus segmentos de fala mais longos na sessão. Os segmentos mais curtos não foram usados, de forma a evitar uma variabilidade excessiva nas características da fala. Consequentemente, o subconjunto *GoN-20seconds* contém 86 ficheiros, com uma duração aproximada de 20 segundos.
- *GoN-10seconds*: este subconjunto foi construído com base na divisão de cada ficheiro no subconjunto *GoN-20seconds* em aproximadamente 2 metades, resultando um subconjunto de 172 ficheiros com uma duração aproximada de 10 segundos cada.

O uso destes três subconjuntos com ficheiros de dimensões diferentes permite efetuar uma análise do efeito desta variável na classificação da personalidade. O nosso objetivo é verificar se a expressão dos traços de personalidade na fala se associa a estruturas linguísticas com uma determinada duração temporal e, portanto, se existem limitações no uso de unidades muito curtas para a classificação da personalidade. A Tabela 2 ilustra o número de exemplos (ficheiros) em cada classe (nível alto ou baixo para um traço) para os três subconjuntos do *corpus* GoN:

	GoN-10seconds					*GoN-20seconds*					*GoN-complete*				
Traço	O	C	E	A	N	O	C	E	A	N	O	C	E	A	N
#Altos	80	96	112	112	76	40	48	56	56	38	10	12	14	14	10
#Baixos	92	76	60	60	96	46	38	30	30	48	12	10	8	8	12

Tabela 2: Número de exemplos em cada classe (nível alto/baixo para um traço) para os três subconjuntos do *corpus* GoN.

As gravações de vídeo originais no *corpus* GoN foram anotadas em termos dos traços de personalidade *Big-Five* por três avaliadores experientes (uma psicóloga e duas peritas em processamento de fala), com

recurso ao questionário de personalidade BFI-10. Essas anotações foram usadas como referência neste trabalho. As medidas da concordância entre as distintas anotadoras em termos do coeficiente Fleiss' Kappa são 0,67 para abertura, 0,15 para conscienciosidade, 0,29 para extroversão, 0,07 para amabilidade e 0,21 para neuroticismo (valor médio de 0,28). Embora não sejam diretamente comparáveis, devido às diferentes configurações experimentais, esses valores estão de acordo com os descritos na literatura, por exemplo, em John & Robins (1993), visto tratar-se de uma tarefa bastante complexa, baseada apenas em propriedades acústicas de ficheiros com reduzida dimensão.

2.2. Pistas acústico-prosódicas para a classificação automática de personalidade

As experiências realizadas neste trabalho utilizam dois conjuntos de características de referência, frequentemente referidas na literatura, extraídas com a ferramenta de acesso público openSMILE (Eyben et al., 2013), juntamente com um conjunto de pistas alicerçadas em conhecimento linguístico prévio (Batista et al., 2012).

2.2.1. Pistas de referência em tarefas paralinguísticas

O primeiro conjunto de referência (IS2012) foi criado no escopo do *Interspeech 2012 Speaker Trait Challenge* como um conjunto exaustivo e não específico de pistas acústico-prosódicas para tarefas nas áreas da paralinguística e de processamento de fala em geral. Este conjunto consiste em 6125 características obtidas a partir de 64 descritores de baixo nível relativos à energia (e.g., *loudness*, energia RMS, *zero-crossing rate*), ao espetro (e.g., coeficientes RASTA e MFCC, fluxo espetral), à frequência fundamental e à qualidade da voz (e.g., *harmonics-to-noise ratio*, *jitter*, *shimmer*). Sobre estes descritores são aplicadas numerosas medidas estatísticas (*functionals*) ao nível do ficheiro (por exemplo, valor médio, desvio padrão, percentis, etc.) de forma a obter as 6125 características finais (veja-se Schuller et al. (2015), tabelas 1 e 2, para uma enumeração exaustiva dos descritores de baixo nível e das medidas estatísticas usados na obtenção do conjunto IS2012). Também é usado como referência o conjunto de pistas eGeMAPS (Eyben et al., 2016), um subconjunto mais específico que consiste em 88 características relativas à energia, ao espetro, à frequência fundamental e à qualidade da voz. Este conjunto é bem conhecido pela sua utilidade numa ampla gama de tarefas paralinguísticas, sobretudo associadas à deteção de emoções.

2.2.2. Pistas baseadas em conhecimento linguístico prévio

O recurso a pistas baseadas em conhecimento linguístico (*knowledge-based features* ou *KB-features*) é motivado por duas razões. Em primeiro lugar, a necessidade de continuar o estudo e desenvolvimento de conjuntos compactos de pistas acústico-prosódicas projetadas especificamente para tarefas de paralinguística computacional (no nosso caso, classificação de personalidade), em contraste com outras abordagens genéricas que têm sido predominantes nos últimos anos, como o conjunto IS2012 que foi empregue no *Interspeech 2012 Speaker Trait Challenge*. Em segundo lugar, as pistas usadas tradicionalmente na literatura são geralmente calculadas aplicando diferentes medidas estatísticas ao nível do ficheiro sobre descritores de baixo nível extraídos com janelas fixas de milissegundos. Ao aplicar janelas de análise de milissegundos, grande parte da informação de fala presente na unidade de fala é perdida. A informação sobre as estruturas temporais internas e a dinâmica de fala é crucial na classificação da personalidade dos falantes, uma vez que os traços de personalidade são mais constantes do que as emoções, por exemplo, sendo que as janelas de análise devem ser mais alargadas.

As pistas baseadas em conhecimento linguístico partem de um alinhamento fonético preliminar dos ficheiros de fala, realizado com base em modelos acústicos do reconhecedor de fala AUDIMUS (Meinedo, Viveiros & Neto, 2008). Nas nossas experiências, usámos modelos de fones em francês e português para efetuar o alinhamento inicial. Tais alinhamentos são a base para extrair pistas relacionadas com todos os

parâmetros prosódicos (duração, energia, frequência fundamental (f_0) e voz). Também permitem caracterizar cada segmento de fala com recurso a *n-gramas* de fones (ou seja, sequências de *n* fones). Com base nesses alinhamentos, extrai-se também a informação de *inter pausal units* (IPUs), que são sequências de fones delimitadas por silêncios que correspondem a unidades prosódicas. Essas unidades permitem-nos calcular medidas estatísticas e características intra e inter-IPUs, características baseadas em durações dos fones e das IPUs, e características prosódicas a partir das sequências de IPUs a um nível macro-estrutural.

As experiências apresentadas neste trabalho utilizaram um conjunto de 41 pistas baseadas em conhecimento linguístico prévio, incluindo a duração da fala com/sem silêncios internos, medidas de ritmo, como as taxas de fala e articulação (número de fones ou sílabas dividido pela duração da fala com e sem silêncios internos, respetivamente), e a taxa de fonação (duração da fala sem silêncios internos dividida pela duração da fala incluindo silêncios internos). Outras características consistem em medidas estatísticas da f_0, energia, *jitter* e *shimmer*, como média, mediana, desvio padrão, dinâmica, amplitude de movimento e declive da frequência fundamental e da energia, calculados intra e inter-IPUs (Batista et al., 2012). As pistas de f_0 são normalizadas numa escala de semitons. Para além das pistas descritas, extraímos também características prosódicas ao nível de unidades similares a frases, essencialmente mais elaboradas e que evolvem as sequências de IPUs, expressas em termos de desvio padrão, declive e concavidade inter-IPUs. O programa Snack Sound Toolkit[2] foi usado para extrair a frequência fundamental e a energia do sinal de fala. O *jitter* e o *shimmer* foram extraídos dos descritores de baixo nível do openSMILE. As pistas baseadas em conhecimento linguístico ainda não tinham sido aplicadas em tarefas de classificação de personalidade, tendo sido combinadas com as pistas eGeMAPS, para obter um melhor desempenho, perfazendo, assim, 129 pistas acústico-prosódicas. A combinatória de pistas também se justifica uma vez que as pistas baseadas em conhecimento linguístico complementam a informação presente nas eGeMAPS, levando a um melhor desempenho na classificação de determinados traços de personalidade (*vide* Secção 3.1).

2.3. Modelos para a classificação de personalidade

Como dito anteriormente, os *corpora* de fala anotados em termos de personalidade são escassos e geralmente de reduzida dimensão, o que inviabiliza o uso de metodologias de aprendizagem automática muito elaboradas para aprender modelos complexos. Em geral, os modelos complexos são descritos de acordo com um número alargado de parâmetros que requerem grandes conjuntos de dados de treino. Por essa razão, usámos *support vector machines* (SVM) lineares como modelos estatísticos (classificadores binários) para a classificação automática de personalidade. Os classificadores SVMs lineares são robustos e, mais importante para o estudo, permitem uma interpretação mais acessível do mapeamento entre características de entrada e etiquetas de saída do classificador. Todos esses modelos foram treinados usando a ferramenta WEKA *data mining software*[3]. Foi dada especial atenção à normalização das pistas, dadas as propriedades heterogéneas dos diferentes *corpora* empregues neste trabalho. Os dados foram normalizados atendendo à amplitude de variação [0,1] ou à média zero e variância unitária, como habitual na literatura.

2.3.1. Aprendizagem supervisionada

Como é usado na literatura, cada dimensão de personalidade do modelo *Big-Five* (OCEAN) é considerada neste trabalho como um problema de classificação binária independente. Assim, foram treinados cinco modelos diferentes, correspondendo à abertura, conscienciosidade, extroversão, amabilidade e neuroticismo. Cada modelo SVM é treinado usando dados do *corpus* SPC para atribuir um nível alto/baixo no traço correspondente (denotado como O/NO, C/NC, E/NE, A/NA e N/NN, respetivamente) para cada ficheiro

[2] http://www.speech.kth.se/snack/
[3] https://www.cs.waikato.ac.nz/~ml/index.html

de fala. Finalmente, os desempenhos dos modelos de personalidade são avaliados em quatro conjuntos de teste diferentes (subconjunto de teste do SPC, e conjuntos *GoN-complete*, *GoN-20seconds* e *GoN-10seconds*), para avaliar os modelos em condições homogéneas e heterogéneas (disparidade de idade e língua). Os desempenhos dos modelos são apresentados em termos da *unweighted average recall* (UAR). A UAR é uma medida mais justa e informativa do desempenho de um sistema em contextos em que os dados utilizados para avaliação apresentam um desequilíbrio substancial no número de exemplos para as duas classes em competição (nível alto/baixo num dado traço):

$$(1) \quad \text{UAR} = 0{,}5\left(\frac{\text{TH}}{\text{TH} + \text{FL}} + \frac{\text{TL}}{\text{TL} + \text{FH}}\right)$$

Nesta expressão, TH e FL correspondem a "altos verdadeiros" e "baixos falsos", respetivamente, isto é, o número de ficheiros de teste originalmente anotados com um nível alto num traço que são classificados correta/incorretamente pelo modelo de personalidade, respetivamente; e TL e FH os "baixos verdadeiros" e "altos falsos", isto é, o número de ficheiros originalmente anotados com um nível baixo num traço que são classificados correta/incorretamente pelo modelo, respetivamente. A UAR é expressa em valores entre 0 e 1, ou como uma percentagem entre 0 e 100 (esta última correspondendo ao desempenho ótimo).

2.3.2. Aprendizagem semissupervisionada

A partir dos modelos preliminares descritos na secção anterior, aplica-se um procedimento iterativo de auto-aprendizagem (semissupervisionado). Em cada iteração, o modelo atual é usado para classificar as amostras restantes no conjunto de dados não anotado CNG, e as 100 amostras com as probabilidades de saída máximas para cada classe (nível alto e baixo num determinado traço) são extraídas do conjunto de dados CNG e adicionadas ao conjunto de treino atual, formando um novo conjunto de treino para ser usado na próxima iteração. As etiquetas atribuídas pelo modelo atual são usadas como etiquetas de referência para treinar o modelo na próxima iteração.

3. Resultados

Esta secção é dedicada à apresentação e discussão dos resultados experimentais para a classificação automática da personalidade de crianças portuguesas obtida pelos sistemas descritos acima. São analisados cinco aspetos: 1) o desempenho dos três conjuntos de pistas acústico-prosódicas; 2) o impacto da extensão dos ficheiros; 3) a viabilidade do uso de *corpora* de fala com disparidade de língua e idade; 4) o potencial de usar conjuntos de dados parcialmente não anotados numa abordagem de aprendizagem semissupervisionada, para aprender modelos de personalidade mais robustos; e 5) a relevância de determinadas pistas acústico-prosódicas para a classificação de cada traço de personalidade. Como mencionado na Secção 2.3, duas técnicas diferentes foram usadas para normalizar os vetores de pistas acústico-prosódicas, i.e., normalizações da amplitude [0,1], ou de média zero e variância unitária. Esta etapa de pré-processamento é de suma importância, dadas as características marcadamente diferentes dos três *corpora* usados neste trabalho, e tem um impacto notório nas experiências de aprendizagem semissupervisionada (*vide* Secção 3.2).

3.1. Aprendizagem supervisionada

Para o trabalho apresentado nesta subsecção, adotou-se a mesma configuração experimental empregue no *Interspeech 2012 Speaker Trait Challenge-Personality Sub-challenge* (Schuller et al., 2012; Schuller et al., 2015). As Figuras 1, 2, 3 e 4 apresentam os resultados para os modelos iniciais aprendidos através de uma abordagem supervisionada, conforme descrito na Secção 2.3.1. Apresentamos os resultados sobre os

conjuntos de avaliação SPC e GoN, usando as pistas do IS2012 (6125) e eGeMAPS (88), bem como o conjunto eGeMAPS+*KB-features* (129).

A Figura 1 mostra os resultados (em termos de UAR) obtidos no subconjunto de teste do SPC, ou seja, o desempenho dos modelos de personalidade no caso homogéneo em que os modelos são treinados e avaliados nas mesmas condições (fala de adultos francófonos). **Os resultados mostram que a conscienciosidade, a extroversão e o neuroticismo (em menor escala) podem ser facilmente classificados no *corpus* SPC**. No geral, o conjunto de pistas do IS2012 supera os outros dois conjuntos, exceto para neuroticismo, para o qual a combinação das *KB-features* e as eGeMAPS atinge o melhor resultado. Além disso, o conjunto eGeMAPS+*KB-features* obteve um melhor desempenho do que as eGeMAPS *per se* para extroversão, amabilidade e neuroticismo.

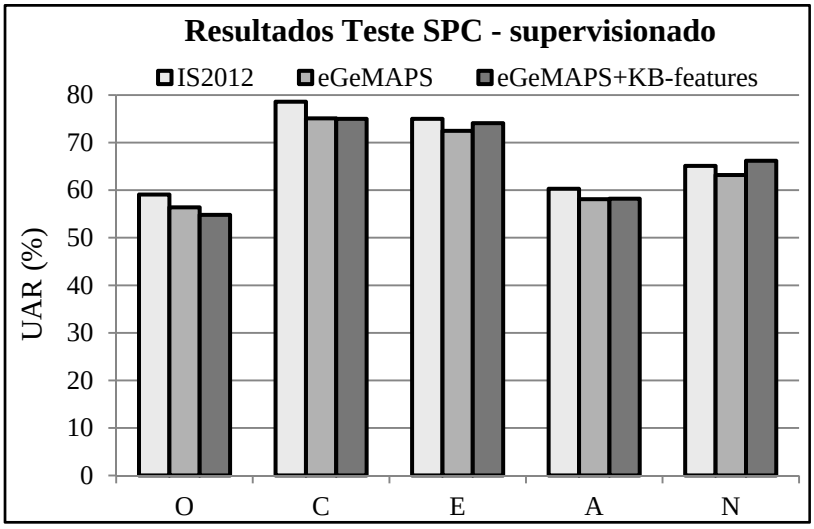

Figura 1: Resultados obtidos no conjunto de dados SPC - modelos iniciais

As Figuras 2, 3 e 4 apresentam os resultados obtidos nos subconjuntos *GoN-complete* (22 ficheiros de fala), *GoN-20seconds* (86 ficheiros) e *GoN-10seconds* (172 ficheiros), respetivamente. Estes resultados correspondem ao caso heterogéneo, no qual os modelos acima mencionados treinados com fala de adultos francófonos (*corpus* SPC) são usados para classificar traços de personalidade na fala de crianças portuguesas (subconjuntos GoN). No geral, estas figuras mostram taxas de desempenho razoáveis para a classificação da abertura, extroversão e amabilidade. Em comparação com os resultados no conjunto de teste do SPC, observamos que os traços de abertura e amabilidade podem ser classificados agora de forma satisfatória nos três subconjuntos do *corpus* GoN. A conscienciosidade, pelo contrário, não é classificada de forma adequada. As razões que motivam as diferenças na classificação dos distintos traços nos conjuntos de avaliação do SPC e GoN estão associadas às características específicas destes dois *corpora*. Por conseguinte, o SPC consiste em gravações de locutores de notícias e comentaristas de rádio adultos, favorecendo a exibição de traços de personalidade como a conscienciosidade e a extroversão e, em alguns casos, o neuroticismo. No *corpus* das crianças, diferentes díades de colegas de turma jogam um jogo cooperativo à procura de uma recompensa final. A especificidade do jogo em questão favorece a expressão de traços mais relacionados com a interação afável entre colegas, tais como a abertura, extroversão e amabilidade, pelo que a conscienciosidade e o neuroticismo não são tão evidentes.

Os resultados para o *Game-of-Nines* não mostram destrinça clara entre os três conjuntos de pistas usados neste estudo. No geral, os melhores resultados são obtidos novamente pelos conjuntos de pistas IS2012 e eGeMAPS+*KB-features*. **Na maioria dos casos, a adição das pistas baseadas em conhecimento ao conjunto das eGeMAPS resulta num melhor desempenho na classificação da personalidade**. Este facto indica que as eGeMAPS não são só por si passíveis de classificar os traços de personalidade, uma vez que as pistas baseadas em conhecimento prévio parecem ser capazes de modelar certas pistas acústico-prosódicas não presentes nos outros dois conjuntos, como as distintas medidas extraídas intra- e inter-IPUs e a dinâmica de variação entre as referidas unidades.

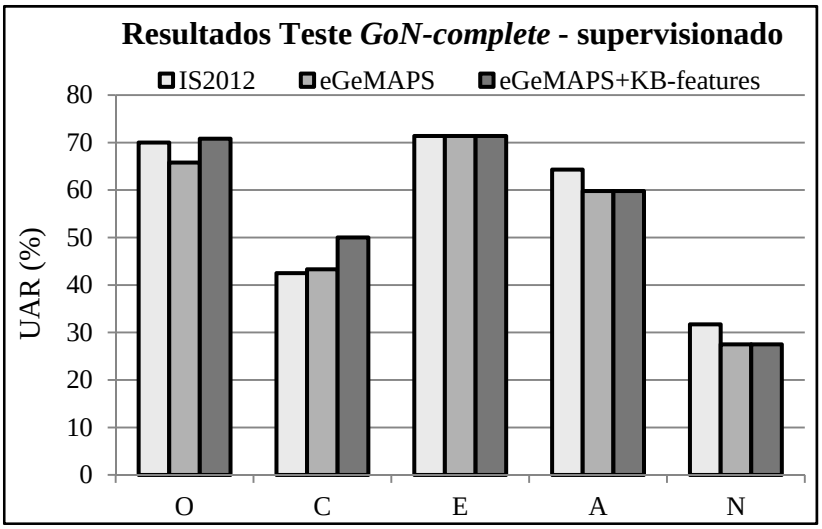

Figura 3: Resultados obtidos no conjunto de dados *GoN-complete* - modelos iniciais

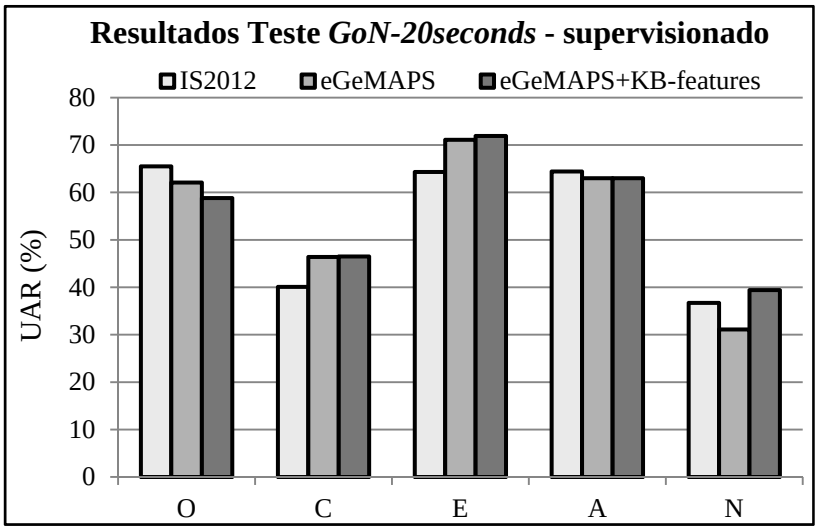

Figura 2: Resultados obtidos no conjunto de dados *GoN-20seconds* - modelos iniciais

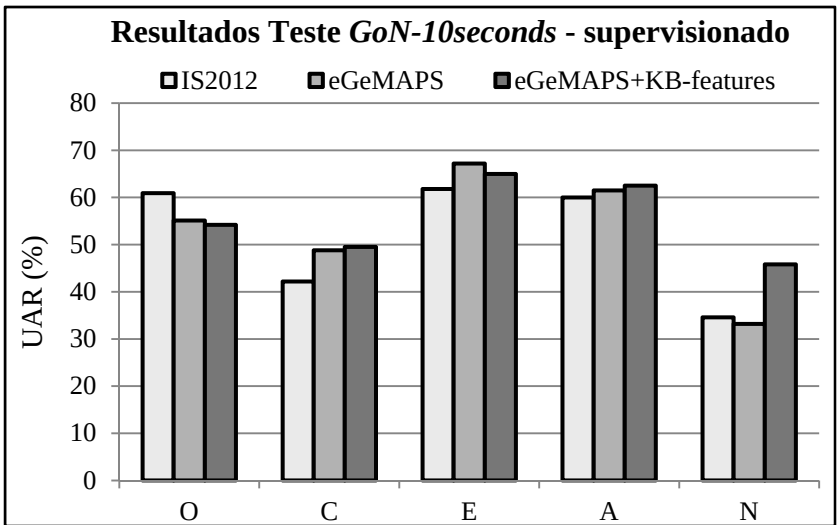

Figura 4: Resultados obtidos no conjunto de dados *GoN-10seconds* - modelos iniciais

Em relação à duração dos ficheiros de fala, os resultados apresentados nas Figuras 2, 3 e 4 mostram que este é um fator que tem impacto na classificação da abertura e da extroversão, apresentando um desempenho decrescente relativamente aos ficheiros mais curtos. Em particular, podemos observar uma deterioração percetível do desempenho à medida que passamos dos ficheiros de 20 segundos para os ficheiros de 10 segundos. De realçar e ainda que, ao contrário do conjunto de dados SPC, que consiste em frases completas com uma duração aproximada de 10 segundos, os ficheiros no subconjunto *GoN-10seconds* são formados pela concatenação de segmentos de fala de curta duração procedentes de interações diádicas, que nem sempre correspondem a uma unidade similar a frase. Por outro lado, observa-se nos resultados que a amabilidade é menos afetada pela duração dos ficheiros e pode ser razoavelmente classificada, mesmo em frases muito curtas. Em geral, os melhores resultados são obtidos no subconjunto *GoN-complete*, o que revela algumas limitações no uso de unidades muito curtas para a classificação de traços de personalidade.

Finalmente, a conclusão mais relevante desta subsecção é a de que **são obtidos resultados razoáveis e consistentes em configurações experimentais muito diferentes para a classificação da abertura, extroversão e amabilidade em fala de crianças portuguesas, usando para isso modelos de personalidade treinados com fala de adultos francófonos**. Os valores da UAR para esses traços estão consistentemente acima de 60% na maioria dos casos, com um valor máximo de 70,8% para abertura (subconjunto *GoN-complete*, eGeMAPS+*KB-features*), de 71,9% para extroversão (subconjunto *GoN-20seconds*, eGeMAPS+*KB-features*) e de 64,4% para amabilidade (subconjunto *GoN-complete*, características openSMILE). Esses resultados apontam para a existência de um conjunto similar e estável de características acústico-prosódicas para esses traços, tanto na fala de adultos francófonos, como de crianças portuguesas. Tal constatação valida a abordagem do presente estudo: usar bases de dados de fala com disparidade de língua e idade para a classificação de personalidade, em contextos em que os dados de fala são especialmente escassos. A hipótese de que dados treinados com distintas variáveis de língua e idade permitiriam classificar automaticamente traços de personalidade não tinha sido ainda testada na literatura, pelo que, até onde nos é dado saber, não poderemos ter valores comparativos em circunstâncias similares.

3.2. Aprendizagem semissupervisionada

No trabalho apresentado nesta subsecção, partimos dos modelos preliminares, treinados mediante uma abordagem totalmente supervisionada, conforme descrito na Secção 2.3.1, e posteriormente aplicamos um procedimento iterativo de aprendizagem semissupervisionada, conforme descrito na Secção 2.3.2. Com este procedimento, o conhecimento implícito extraído de fala não anotada de crianças portuguesas é explorado para melhorar os modelos iniciais de personalidade treinados com fala anotada de adultos francófonos.

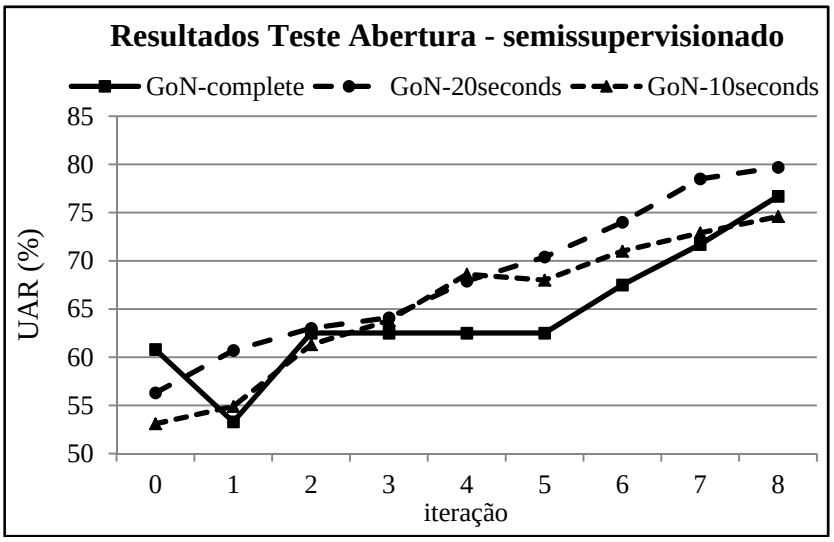

Figura 5: Abertura: resultados obtidos nos conjuntos de dados GoN –aprendizagem semissupervisionada

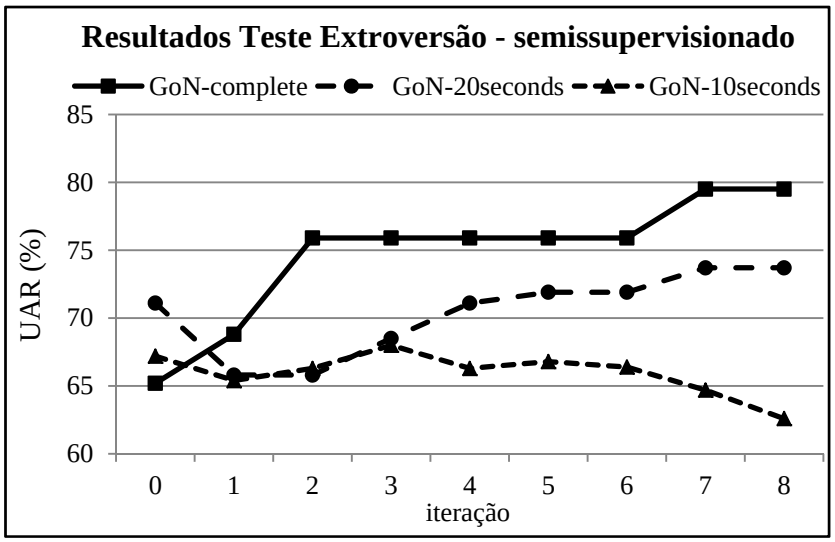

Figura 6: Extroversão: resultados obtidos nos conjuntos de dados GoN –aprendizagem semissupervisionada

As Figuras 5 e 6 mostram os resultados (UAR) obtidos pelos modelos treinados em cada iteração da abordagem semissupervisionada para os traços abertura e extroversão, respetivamente. São apresentados os resultados nos três subconjuntos de teste do *corpus* GoN, usando as características eGeMAPS, sendo o comportamento geral semelhante para os outros dois conjuntos usados neste trabalho. Verifica-se que **a incorporação progressiva de dados não etiquetados nas sucessivas iterações do treino produz, em geral, uma melhoria no desempenho dos sistemas para os traços abertura e extroversão**. Isto é particularmente significativo para o traço abertura, para o qual melhorias absolutas entre 15,9% e 23,4% são alcançadas quando se comparam os modelos finais com os iniciais, independentemente da extensão dos ficheiros. Para a extroversão, é obtida uma melhoria absoluta de 14,3% no subconjunto *GoN-complete*, porém, a adaptação dos modelos iniciais com dados não anotados parece ser contraproducente para os segmentos de fala mais curtos do subconjunto *GoN-10seconds*. Para a amabilidade, a abordagem semissupervisionada não resulta em melhores taxas de classificação de personalidade.

Como nota complementar, as configurações experimentais (isto é, tipo de pistas acústico-prosódicas e método de normalização dos vetores de pistas) que resultaram nos melhores modelos de personalidade nos casos supervisionados e semissupervisionados não foram as mesmas, o que explica o facto de a UAR, para os modelos iniciais nas Figuras 5 e 6 (iteração 0), não corresponder aos melhores resultados apresentados na Secção 3.1. Essa discrepância nas configurações experimentais ótimas é fundamentalmente motivada pelo uso na abordagem semissupervisionada de dois *corpora* de fala com características muito diferentes para treinar e adaptar os modelos de personalidade: a configuração experimental que pode funcionar melhor na abordagem supervisionada (fala de adultos francófonos) pode não ser a melhor na abordagem semissupervisionada (com fala de adultos francófonos e de crianças portuguesas). Em particular, notámos que o método empregue para a normalização das características tem um impacto significativo nos resultados, embora não haja evidência de padrões consistentes para todos os traços de personalidade.

Os resultados preliminares apresentados nesta subsecção apontam para vantagens no uso de dados não anotados como método para superar a falta de dados anotados, como na tarefa de classificação automática da personalidade de crianças descrita neste trabalho.

3.3. Relevância das pistas acústico-prosódicas para a classificação dos traços de personalidade

Apresenta-se nesta secção uma discussão sobre a relevância de determinadas pistas acústico-prosódicas para a classificação de cada traço de personalidade. A relevância das pistas foi extraída com base nos pesos normalizados dos classificadores. A Tabela 3 mostra as dez pistas acústico-prosódicas mais informativas para a classificação de cada um dos cinco traços de personalidade do modelo *Big-Five*, de acordo com os modelos SVM de personalidade treinados usando o *corpus* SPC e o conjunto de eGeMAPS+*KB-features* (Secção 3.1). Restringimos o nosso estudo ao conjunto de características eGeMAPS+*KB-features*, o que permite focar a análise e a discussão num conjunto relativamente compacto de 129 características acústico-prosódicas, conhecidas pela sua informatividade em tarefas de classificação de personalidade. Adicionalmente, seria muito mais difícil extrair conclusões sólidas do conjunto de características IS2012, mais genérico e redundante.

Característica \ Traço de personalidade		O[4]	C	E	A	N
Energia	*loudness_amean*				9+	
	loudness_percentile50.0				4+	
	loudness_percentile80.0				6+	
	loudness_meanFallingSlope	*-		5-	7+	4-
	loudnessPeaksPerSec	*-	1-	1-		1-
	energy.range	*-	8-	10-		
	energy.stdev	*-				
	energy.dynamics	*-	2-	8-		
Informação espectral	*mfcc1_amean*				8-	
	mfcc1V_amean			9+	3-	
	mfcc2V_stddevNorm					5-
	mfcc4V_amean	*-				
	logRelF0-H1-H2_amean		5-			
	hammarbergIndexV_amean				1-	
	alphaRatioV_amean				2+	
	spectralFlux_amean					8-
	spectralFluxUV_amean					9-
Frequência fundamental	*F0semitoneFrom27.5Hz_percentile50.0*				10+	7-
	F0semitoneFrom27.5Hz_percentile80.0			6-		
	pitch_st.dynamics	*-		3-		
	speech.pitch_st.median.dynamics			4-		2+
Informação temporal	***phones.speech-rate***	*-	4-			
	phones.ipu-speech-rate	*-	3-			
	useful_speech		6-	7-		
	phones.silence_ratio		7+			
	ipu.silence_ratio		10+			
	StddevUnvoicedSegmentLenght					10-
Qualidade da voz	*jitterLocal_stddevNorm*	*+	9+			3+
	shimmerLocaldB_stddevNorm			2+	5-	
	HNRdBACF_stddevNorm					6+

Tabela 3: Pistas acústico-prosódicas mais relevantes para a classificação dos traços de personalidade de crianças portuguesas (O: *openness* ou abertura; C: *conscientiousness* ou conscienciosidade; E: *extroversion* ou extroversão; A: *agreeableness* ou amabilidade; N: *neuroticism* ou neuroticismo). O sinal de "+" significa uma correlação positiva entre a pista e o traço de personalidade, enquanto o sinal de "-" significa o inverso.

As características relacionadas com a **dinâmica** e **variações de energia** e as **propriedades da voz** são muito informativas para a classificação da **amabilidade**. Tal como verificado para outras línguas, os traços

[4] Os pesos dos modelos discutidos nesta secção são extraídos dos resumos de saída fornecidos pelo software usado para treinar os modelos de personalidade (WEKA). Devido a limitações na precisão decimal nesses resumos, só foi possível obter valores arredondados dos pesos para o caso da abertura. Embora isso permita identificar as dez características mais relevantes, não é possível estabelecer uma ordenação das pistas, como para os outros traços de personalidade.

mais distintivos da amabilidade compreendem a dinâmica e as variações de energia mais amplas (como manifestado pelo sinal positivo na tabela para todas as pistas associadas à energia) e por propriedades de voz associadas a uma voz suave. As variações mais amplas do parâmetro energia deste traço são contrastivas com todos os outros traços, ao contrário do que é referido na literatura (veja-se para uma revisão o estudo de Mohammadi & Vinciarelli, 2012). Por conseguinte, o traço **extroversão** é também associado a valores amplos de energia e de f_0, porém, nos dados de treino dos adultos franceses, os boletins informativos, a extroversão não se caracteriza por gamas amplas de energia ou e de $f_{0,}$ mas antes por valores estáveis das pistas em questão, uma vez que o carácter informativo do que estão a reportar não se configura nos moldes das dinâmicas representativas de diálogos espontâneos. Na linha do referido para a extroversão, também o traço **abertura** apresenta valores de energia e de f_0 mais baixos, facto que se crê estar novamente associado à natureza dos dados, ou seja, ao carácter informativo e jornalístico.

Tal como descrito na literatura também para o traço **conscienciosidade**, o **baixo rácio de silêncios** é assaz informativo, o que significa que um falante consciencioso (também chamado "competente" por Mohammadi & Vinciarelli, 2012) tende a falar sem produção de silêncios longos. Ao contrário do referido estudo, os falantes caracterizados como conscienciosos tendem a produzir enunciados com menor dinâmica de energia do que um falante caracterizado como amável, tendem a falar mais pausadamente, mas controlando os silêncios para que estes não sejam longos.

Relativamente ao traço **neuroticismo**, um falante caracterizado como neurótico produz enunciados com **menores valores médios de energia e características de voz associadas à laringalização**, pistas que já foram anteriormente descritas para a identificação de stress (Julião et al., 2015).

De salientar é que as características baseadas em conhecimento linguístico prévio são assaz informativas. No geral, 10 das 41 características que compõem este novo conjunto estão entre as mais informativas para a classificação dos traços de personalidade (em negrito na Tabela 3). Estes resultados sugerem os benefícios da utilização de características acústico-prosódicas específicas da tarefa baseadas em conhecimento linguístico prévio como um complemento eficaz ao conjunto de características eGeMAPS usado para a classificação computacional da personalidade.

4. Conclusões

A principal motivação deste trabalho é a necessidade de desenvolver métodos para superar a escassez de dados de fala devidamente anotados para a classificação de personalidade em domínios (língua e população-alvo) com poucos recursos. Com este objetivo, este trabalho investiga uma abordagem de treino semissupervisionado, na qual conjuntos de dados heterogéneos e parcialmente anotados podem ser usados para contornar a escassez de *corpora* de fala para o domínio alvo (crianças portuguesas). Concebemos aqui uma configuração experimental com disparidade de idade e língua, em que modelos de personalidade, treinados usando um conjunto anotado de dados de fala de adultos francófonos, são refinados de forma iterativa com recurso a um conjunto não anotado de dados de fala de crianças portuguesas. Também investigamos o impacto de distintos conjuntos de pistas baseadas em conhecimento linguístico prévio sobre pistas acústico-prosódicas genéricas para tarefas de classificação de personalidade.

O presente trabalho traz uma visão sobre a classificação de personalidade em fala espontânea. Os resultados apresentados na Secção 3.1 apontam para a existência de um conjunto similar e estável de características acústico-prosódicas para a abertura, extroversão e amabilidade na fala de adultos fracófonos e crianças portuguesas. Além disso, os resultados da Secção 3.2 mostram as vantagens da abordagem de treino semissupervisionado para extrair conhecimento implícito de *corpora* extensos não anotados.

Em conclusão, estes resultados revelam os benefícios do uso de bases de dados de fala heterogéneas e parcialmente anotadas em tarefas de classificação de personalidade como um método para superar a escassez

de dados em áreas com poucos recursos. A discussão apresentada na Secção 3.3 salienta a importância de uma seleção adequada de conjuntos de características específicas para a classificação de cada traço de personalidade. Os resultados também sugerem os benefícios do uso de características acústico-prosódicas específicas para a tarefa baseadas em conhecimento prévio como um complemento eficaz para os conjuntos de características tradicionalmente usados para a classificação computacional da personalidade.

Por forma a realizar experiências mais exaustivas e relevantes nesta linha de investigação, será importante adquirir mais dados de fala. Em particular, isso permitirá empregar procedimentos de seleção de características e metodologias de aprendizagem automática mais elaboradas, com o objetivo de desenvolver sistemas automáticos de classificação de personalidade mais robustos.

Referências

Batista, Fernando, Helena Moniz, Isabel Trancoso & Nuno Mamede (2012) Bilingual experiments on automatic recovery of capitalization and punctuation of automatic speech transcripts. *IEEE Transactions on Audio, Speech and Language Processing*, 20 (2), pp. 474–485.

Campos, Joana, Patrícia Alves-Oliveira & Ana Paiva (2016) Looking for conflict: gaze dynamics in a dyadic mixed-motive game. *Autonomous Agents and Multi-Agent Systems*, 30 (1), pp. 112–135.

Caspi, Avshalom, HonaLee Harrington, Barry Milne, James W. Amell, Reremoana F. Theodore, Terrie E. Moffitt (2003) Children's behavioral styles at age 3 are linked to their adult personality traits at age 26. *Journal of Personality* 71 (4), pp. 495–514.

Chapelle, Olivier, Bernhard Schölkopf & Alexander Zien, eds. (2006). *Semi-supervised learning*. Cambridge: Massachusetts Institute of Technology.

Cloninger, Susan (2009) Conceptual issues in personality theory. In Philip J. Corr & Gerald Matthews (eds.) *The Cambridge Handbook of Personality Psychology* (4). Cambridge: Cambridge University Press, pp. 3–26.

Coutinho, Eduardo, Jun Deng & Björn Schuller (2014) Transfer learning emotion manifestation across music and speech. In *Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN)*, pp. 3592–3598.

Deng, Jun, Zixing Zhang & Björn Schuller (2014) Linked source and target domain subspace feature transfer learning–exemplified by speech emotion recognition. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR 2014)*, pp. 761–766.

Eyben, Florian, Felix Weninger, Florian Groß & Björn Schuller (2013) Recent developments in openSMILE, the Munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, pp. 835–838.

Eyben, Florian, Klaus R. Scherer, Björn W. Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y. Devillers, Julien Epps, Petri Laukka, Shrikanth S. Narayanan & Khiet P. Truong (2016) The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing*, 7 (2), pp. 190–202.

Hämäläinen, Annika, Fernando Miguel Pinto, Silvia Rodrigues, Ana Júdice, Sandra Morgado Silva, António Calado, José Miguel Salles Dias (2013) A multimodal educational game for 3-10-year-old children: collecting and automatically recognising european portuguese children's speech. In *Proceedings of Workshop on Speech and Language Technology in Education.*

John, Oliver P. & Richard W. Robins (1993) Determinants of interjudge agreement on personality traits: the big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, 61 (4), pp. 521–551.

Julião, Mariana, Jorge Silva, Ana Aguiar, Helena Moniz & Fernando Batista (2015) Speech features for discriminating stress using branch and bound wrapper search. In José-Luis Sierra-Rodríguez, José-Paulo Leal & Alberto Simões (eds) *Languages, Applications and Technologies. SLATE 2015*. Cham: Springer, Communications in Computer and Information Science, vol 563, pp. 3–14.

Kagan, Jerome (1996) Three pleasing ideas. *American Psychologist* 51, pp. 901–908.

Kelley, Harold H., Linda Linden Beckman & Claude S. Fischer (1967) Negotiating the division of a reward under incomplete information. *Journal of Experimental Social Psychology*, 3 (4), pp. 361–398.

Lewis, Michael (1997) *Altering fate: why the past does not predict the future*. New York: Guilford Press.

Mairesse, François, Marilyn A. Walker, Matthias R. Mehl & Roger K. Moore (2007) Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research* 30 (1), pp. 457–500.

Meinedo, Hugo, Márcio Viveiros & João Neto (2008) Evaluation of a Live Broadcast News Subtitling System for Portuguese. In *Proceedings of Interspeech 2008*, pp. 508–511.

Mohammadi, Gelareh & Alessandro Vinciarelli (2012) Automatic personality perception: prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing* 3 (3), pp. 273–284.

Polzehl, Tim, Sebastian Möller & Florian Metze (2010a) Automatically assessing acoustic manifestations of personality in speech. In *Proceedings of the Spoken Language Technology Workshop*, pp. 7–12.

Polzehl, Tim, Sebastian Möller & Florian Metze (2010b) Automatically assessing personality from speech. In *Proceedings of the International Conference on Semantic Computing*, pp. 134–140.

Polzehl, Tim, Sebastian Möller & Florian Metze (2011) Modeling speaker personality using voice. In *Proceedings of Interspeech 2011*, pp. 2369–2372.

Rammstedt, Beatrice & Oliver P. John (2007) Measuring personality in one minute or less: a 10-item short version of the Big Five inventory in English and German. *Journal of Research in Personality* 41, pp. 203–212.

Roberts, Brent W., Kate E. Walton, Wolfgang Viechtbauer (2006) Patterns of mean-level change in personality traits across the life course: a meta-analysis of longitudinal studies. *Psychological Bulletin* 132 (1), pp. 1–25.

Schuller, Björn W., Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi & Benjamin Weiss (2012) The Interspeech 2012 speaker trait challenge. In *Proceedings of Interspeech 2012*, pp. 254–257.

Schuller, Björn W., Stefan Steidl, Anton Batliner, Elmar Nöth, Alessandro Vinciarelli, Felix Burkhardt, Rob van Son, Felix Weninger, Florian Eyben, Tobias Bocklet, Gelareh Mohammadi & Benjamin Weiss (2015) A survey on perceived speaker traits: personality, likability, pathology, and the first challenge. *Computer Speech & Language* 29 (1), pp. 100–131.

Solera-Ureña, Rubén, Helena Moniz, Fernando Batista, Ramón Fernández-Astudillo, Joana Campos, Ana Paiva & Isabel Trancoso (2016) Acoustic-prosodic automatic personality trait assessment for adults and children. In Alberto Abad, Alfonso Ortega, António Teixeira, Carmen García Mateo, Carlos D. Martínez Hinarejos, Fernando Perdigão, Fernando Batista & Nuno Mamede (eds.) *Advances in Speech and Language Technologies for Iberian Languages: Third International Conference, IberSPEECH 2016*. Cham: Springer International Publishing AG, pp. 192–201.

Solera-Ureña, Rubén, Helena Moniz, Fernando Batista, Vera Cabarrão, Anna Pompili, Ramón Fernández Astudillo, Joana Campos, Ana Paiva & Isabel Trancoso (2017) A semi-supervised learning approach for acoustic-prosodic personality perception in under-resourced domains. In *Proceedings of Interspeech 2017*, pp. 929–933.

Vinciarelli, Alessandro & Gelareh Mohammadi (2014) A survey of personality computing. *IEEE Transaction on Affective Computing* 5 (3), pp. 273–291.