



# Previsão de Comunicação Eficácia durante as habilidades de mídia Treinamento Usando Comercial Reconhecimento não verbal automático Sistemas

Monica Pereira<sup>1\*</sup>, Hongying Meng<sup>2</sup> e Kate Hone<sup>3</sup>

<sup>1</sup> Departamento de Psicologia, Escola de Ciências Sociais, London Metropolitan University, Londres, Reino Unido,

<sup>2</sup> Departamento de Engenharia Eletrônica e de Computação, Faculdade de Engenharia, Design e Ciências Físicas, Brunel

University London, Londres, Reino Unido, e Ciências <sup>3</sup> Departamento de Ciência da Computação, Faculdade de Engenharia, Design Físicas, Brunel University London, Londres, Reino Unido

## OPEN ACCESS

### Editado por:

Maurizio Codispoti,  
Universidade de Bolonha, Itália

### Revisados pela:

Marcelo Mortilano,  
Université de Genève, Suíça  
Laszlo Hunyadi,  
Universidade de Debrecen, Hungria

### \*Correspondência:

Mônica Pereira  
monica.pereira@londonmet.ac.uk

### Seção de especialidade:

Este artigo foi submetido a  
Ciência da Emoção,  
uma seção da revista  
Fronteiras da Psicologia

**Recebido:** 03 de março de 2021

**Aceito:** 08 de julho de 2021

**Publicado:** 29 de setembro de 2021

### Citação:

Pereira M, Meng H e Hone K (2021)  
Previsão da eficácia da comunicação durante  
o treinamento de habilidades de mídia  
usando sistemas comerciais automáticos de  
reconhecimento não verbal.  
Frente. Psicol. 12:675721.  
doi: 10.3389/fpsyg.2021.675721

É bem reconhecido que os sinais sociais desempenham um papel importante na eficácia da comunicação. A observação de vídeos para entender o comportamento não verbal é o tempo

consome e limita o potencial de incorporar feedback detalhado e preciso desse comportamento em aplicações práticas, como treinamento de habilidades de comunicação ou avaliação de desempenho. O objetivo da pesquisa atual é duplo: (1) investigar se a tecnologia de reconhecimento de emoções pode detectar sinais sociais em entrevistas na mídia e (2) identificar quais combinações de sinais sociais são mais promissoras para avaliar o desempenho dos estagiários em uma entrevista na mídia. Para investigar isso, sinais não verbais foram automaticamente reconhecidos a partir de entrevistas de mídia práticas na câmera conduzidas em um ambiente de treinamento de mídia com um tamanho de amostra de 34. A detecção automatizada de sinais não verbais consiste em recursos multimodais, incluindo expressão facial, gestos de mão, comportamento vocal e 'sinais honestos'. As entrevistas na câmera foram categorizadas em exemplos de comunicação eficazes e ruins com base nas classificações de habilidades de comunicação fornecidas por treinadores e observadores neutros que serviram como uma verdade básica.

Um método de seleção de recursos baseado em correlação foi usado para selecionar sinais associados ao desempenho. Para avaliar a precisão dos recursos selecionados, várias técnicas de classificação de aprendizado de máquina foram usadas. A análise Naive Bayes produziu os melhores resultados com uma medida F de 0,76 e precisão de previsão de 78%. Os resultados revelaram que uma combinação de movimentos corporais, movimentos das mãos e expressão facial são relevantes para estabelecer a eficácia da comunicação no contexto de entrevistas na mídia.

Os resultados do estudo atual têm implicações para a avaliação automática de entrevistas de mídia com uma série de áreas de aplicação potencial, incluindo aprimoramento do treinamento de comunicação, incluindo treinamento de habilidades de mídia atual.

**Palavras-chave:** detecção de sinais sociais, tecnologias comerciais, habilidades de comunicação, treinamento, sinais não verbais, entrevistas na mídia, fusão multimodal

## INTRODUÇÃO

Comunicação hábil em entrevistas de mídia é importante em uma variedade de organizações e funções de trabalho. Investimentos significativos em treinamento são feitos para melhorar as habilidades de comunicação, para que os funcionários relevantes apareçam positivamente na mídia. A comunicação é um fenômeno complexo que é definido como a transmissão de informações de uma pessoa para outra (Fiske, 2010; Knapp et al., 2013; Deveugele, 2015). Pesquisas iniciais em psicologia sugeriram que a comunicação verbal representa apenas 7% da percepção social (Vinciarelli et al., 2009).

No entanto, o peso das mensagens depende do contexto e do tipo de interação social. Portanto, é importante que observações precisas e objetivas de pistas não verbais sejam incorporadas à avaliação do desempenho da mídia e às intervenções de treinamento para melhorar o desempenho. No entanto, as ferramentas atuais para dar suporte a isso são limitadas.

Pesquisas anteriores no campo da análise não verbal baseavam-se apenas na observação metódica e na análise de dados de vídeo, como assistir a horas de vídeos gravados para interpretar situações sociais (Vrij et al., 2000; Hart et al., 2016). Este método de análise do comportamento é sujeito à subjetividade, é demorado e não se adapta a grandes quantidades de dados. Neste artigo, propomos uma alternativa escalável que possibilita uma avaliação de emoções mais rápida, de fácil acesso aos pesquisadores e de mensuração mais objetiva de sinais não verbais para profissionais. Especificamente, exploramos o potencial de uma variedade de tecnologias comerciais de prateleira, reconhecendo uma variedade de sinais não-verbais, para identificar o desempenho hábil no contexto de entrevistas na mídia. As tecnologias comerciais disponíveis no mercado foram propostas como um meio eficaz de detectar sinais não verbais na natureza (Dupré et al., 2018; Pereira e Hone, 2021).

Os pesquisadores optaram por usar produtos comerciais de prateleira tecnologias em vez de desenvolver soluções sob medida, a fim de fornecer uma prova de conceito relativamente rápida para a relevância de uma variedade de canais na avaliação do desempenho das habilidades de mídia. Essa abordagem também foi adotada para permitir a transferência rápida para os usuários finais, uma vez que as tecnologias potenciais já podem ser obtidas comercialmente. A lógica é ajudar a reduzir o espaço de design para futuras soluções sob medida. Além disso, o foco está na aplicabilidade funcional de soluções desenvolvidas com tecnologia de afeto. Isso pode ser benéfico, pois essa tecnologia permite que o usuário/pesquisador processe gravações/imagens localmente, onde a classificação de emoções e expressões são extraídas e produzidas pelo classificador do software (Dupré et al., 2018).

No estudo atual, os participantes que participaram de entrevistas durante o treinamento de habilidades de mídia foram gravados. Os dados foram recolhidos utilizando várias tecnologias para permitir a detecção de emoções e comportamentos não verbais. As entrevistas também foram avaliadas quanto à qualidade das habilidades de comunicação por treinadores e observadores neutros usando um instrumento de pesquisa padronizado. A análise concentrou-se em explorar quais dos sinais detectados estavam associados a um desempenho bom ou ruim, conforme avaliado por observadores humanos, e foram tiradas conclusões sobre o potencial uso futuro de tais tecnologias.

Para o conhecimento do pesquisador, não houve estudos investigando se a tecnologia comercial pode detectar

sinais sociais multimodais para comunicação eficaz e nenhum estudo investigando a comunicação no contexto de entrevistas na mídia. Portanto, o objetivo deste duplo: **(1) investigar se a tecnologia comercial automatizada de reconhecimento de afeto pode detectar sinais não-verbais em uma interação diádica e (2) investigar qual combinação de sinais multimodais é necessária para uma comunicação eficaz em uma entrevista na mídia.**

Portanto, a questão de pesquisa (RQ) deve ser investigada: Qual combinação de sinais não verbais é necessária para avaliar o desempenho das habilidades de comunicação durante uma entrevista na mídia?

O RQ é explorado usando os dados de uma variedade de entrevistas práticas com a mídia durante os workshops de treinamento de mídia. A pesquisa atual fornece quatro contribuições principais:

- (1) Fornece uma compreensão mais profunda da comunicação

Habilidades,

- (2) Fornece evidências de que o uso desse tipo de tecnologia automatizada pode ser usado para detectar sinais sociais e não-verbais em um contexto pessoa-pessoa, (3) Identifica os sinais relevantes para entrevistas na mídia, (4) Auxilia os treinadores em escolher o melhor tipo de tecnologia para usar no treinamento para melhorar os resultados de desempenho.

## FUNDO

Nesta seção, revisamos brevemente trabalhos anteriores sobre sinais não-verbais antes de considerar o papel dos sinais não-verbais no caso específico de entrevistas na mídia, que constituem o foco deste artigo. Em seguida, apresentamos brevemente a pesquisa usando detecção automatizada de sinais não verbais e limitações em pesquisas anteriores; por fim, apresentamos ao leitor as finalidades e objetivos da presente pesquisa.

### Comunicação não-verbal

As complexidades da comunicação residem nas funções do contexto e relacionamento. Para entender a comunicação, é preciso reconhecer que a comunicação é multimodal (Hunyadi, 2019). Há um grande corpo de evidências de que os sinais não verbais são importantes em muitos tipos de interação humana (ver Knapp et al., 2013). Estudos de sinais não verbais mostram que a comunicação é tipicamente caracterizada pela interação complexa de sinais recíprocos entre interlocutores (Knapp et al., 2013). Pensa-se que vários sinais não verbais correspondem a emoções sentidas internamente que são expressas consciente ou inconscientemente. Em termos evolutivos, exibir emoções beneficia tanto os emissores quanto os receptores nas interações sociais. Esses sinais são comunicados por vários canais; como expressões faciais, comportamento vocal (ou seja, tom de voz e rajadas vocais), gestos e postura (Adams e Kveraga, 2015).

O rosto humano contém uma infinidade de funções diferentes.

Uma dessas funções é expressar emoções. Desde o trabalho inicial de Darwin (2015) até o trabalho empírico posterior de Ekman et al. (1969) e Ekman e Friesen (1971) houve inúmeras sugestões para a existência de universal (reconhecido em

todas as culturas) emoções básicas que são exibidas em expressões faciais reconhecíveis.

Ekman identificou seis emoções básicas; raiva, medo, nojo, felicidade, tristeza e surpresa e sete expressões faciais universalmente reconhecidas, abrangendo desprezo, bem como as seis emoções básicas (Ekman e Friesen, 1986). Outras teorias propuseram várias outras emoções básicas; exemplos incluem ansiedade, vergonha e prazer (Ortony e Turner, 1990). No entanto, essa teoria tem sido considerada reducionista e simplista (Gross e Feldman-Barrett, 2011). No entanto, a pesquisa em emoção continua a aplicar essa teoria (ver Ekman, 2016).

Ekman (1997) desenvolveu um sistema manual para rotular ações faciais. Este sistema é chamado de Facial Action Coding System (FACS). Este sistema é baseado no mapeamento dos músculos da face para diferentes expressões faciais e define um total de 18 Unidades de Ação (AUs) na face inferior, 9 na face superior, 9 para a posição dos olhos, 11 para a posição da cabeça e 14 movimentos diversos.

Os codificadores humanos usam esse sistema para codificar manualmente todas as expressões faciais. Como esses AUs são independentes de interpretações, eles podem ser usados no reconhecimento de emoções básicas (EMFACS). Por exemplo, as AUs envolvidas em uma demonstração emocional de felicidade são as Unidades de Ação 6 (Levantar a bochecha) e 12 (Puxar o canto do lábio).

O comportamento vocal não-verbal contém todas as pistas que cercam as mensagens verbais que influenciam o significado do conteúdo falado. Existem cinco componentes principais no comportamento vocal não-verbal, que incluem vocalizações linguísticas, vocalizações não linguísticas, qualidade de voz, silêncio e troca de turnos. Cada um deles contribui para a percepção social de uma mensagem (Hall et al., 2019). Por exemplo, a entonação vocal pode alterar o tom de uma mensagem para ser irônico ou sarcástico.

A qualidade da voz está relacionada com as características prosódicas, como tom, energia e tempo. Isso explica como algo é dito. Ele transmite emoções como raiva ou medo. Essas duas emoções são exibidas gritando (Lieberman, 1976). O pitch influencia a percepção de dominância e extroversão, a fluência relaciona-se com a persuasão (Vinciarelli et al., 2009). Vocalizações linguísticas são não-palavras usadas no lugar de palavras como uhm ou ah ha. Estes são chamados de segregados que são frequentemente usados em situações sociais quando envergonhados ou têm dificuldade com uma interação social (Glass et al., 1982). As vocalizações não linguísticas incluem explosões como choro, gemidos, risos ou soluços. O choro, por exemplo, é muitas vezes associado ao espelhamento (Chartrand e Bargh, 1999), o que aumenta os laços sociais.

Os gestos são frequentemente usados para regular as interações, alterando o movimento do braço, posturas e cinemática para exibir emoções (Pollick et al., 2001; Gross et al., 2012). Por exemplo, polegar para cima para indicar reconhecimento (Altman, 1978). Os gestos também podem ser usados para exibir informações inconscientes, como o uso de adaptadores, como dobrar os braços ou mover as pernas ritmicamente para indicar tédio (Pentland e Heibeck, 2010). As posturas também são assumidas consciente ou inconscientemente, pois tendem a revelar as atitudes das pessoas em relação a uma situação social (Schefflen, 1964).

Sabe-se que a comunicação entre dois interlocutores depende do objetivo e do contexto. Por exemplo, sinais não verbais que foram detectados e identificados como potencialmente importantes em uma entrevista de emprego são sorrir mais (Naim et al., 2016), enquanto em um ambiente de saúde a tomada de turnos, a proporção da fala,

volume, tom, sorriso, franzir a testa, inclinar a cabeça, balançar a cabeça, tremer e movimentos gerais do corpo foram extraídos (Liu et al., 2016).

Na sala de aula, as pistas não verbais extraídas durante as apresentações foram a prosódia, a qualidade da voz e a atividade gestual (Cheng et al., 2014). Esses estudos sugerem que exibições apropriadas de sinais não verbais diferem dependendo do contexto.

No geral, existe uma literatura muito rica sobre o papel dos sinais não verbais na comunicação eficaz. Esta seção descreveu brevemente alguns dos principais canais de comunicação não verbal e destacou a importância de observar os sinais no contexto das trocas recíprocas.

O contexto da comunicação também é importante, por isso consideramos agora o contexto de comunicação específico das entrevistas na mídia, que constitui a base deste estudo.

### Comunicação não-verbal em entrevistas na mídia Os manuais

de treinamento em mídia normalmente sugerem alguns

comportamentos específicos que devem ser evitados em entrevistas na mídia.

Comportamentos como falta de convicção vocal, falta de contato visual, velocidade de fala rápida, voz monótona e hesitação são indicativos de nervosismo, incerteza e tédio e influenciam como o entrevistado é percebido pelo público (Taylor, 2015). Um comportamento adicional que pode ser interpretado como tédio são os movimentos excessivos, como balançar e balançar, principalmente quando a outra pessoa está falando (Tao e Tan, 2009).

Os sinais combinatórios provavelmente serão importantes para uma boa entrevista na mídia; como espelhar os movimentos do entrevistador, manter contato visual e sorrir. Juntos, esses sinais sugerem que o entrevistado está ouvindo, sinaliza a troca de turnos na conversa (Ho et al., 2015; Taylor, 2015), ilustra confiança, honestidade e domínio (Knutson, 1996; Lapidot-Lefler e Barak, 2012).

Há um número limitado de estudos que exploraram empiricamente as relações entre comportamentos não-verbais observáveis e julgamentos subjetivos do observador no contexto de entrevistas na mídia. Esses estudos geralmente se concentram em pequenas amostras de entrevistas com entrevistados de alto nível, como políticos. Por exemplo, Babad (1999) correlacionou o julgamento do observador sobre a impressão global (positiva/negativa) criada em uma entrevista na mídia com um conjunto de julgamentos do observador em relação ao comportamento observável. Este artigo continha três estudos que se concentraram nos comportamentos de seis entrevistados que participaram de entrevistas políticas na televisão e encontraram vários padrões comuns entre esses indivíduos. Os comportamentos que pareciam criar uma impressão positiva incluíam sorrir, um rosto relaxado, acenar com a cabeça e fazer movimentos circulares com as mãos. Por outro lado, os comportamentos associados a julgamentos negativos incluíam movimentos de bater com as mãos, inclinar-se para a frente e piscar. Estudos como este têm sido tipicamente de pequena escala devido ao desafio de codificar manualmente os comportamentos comunicativos não-verbais em estudo. No entanto, o desenvolvimento de tecnologias para detectar automaticamente sinais não verbais apresenta maior oportunidade de desenvolver uma compreensão das pistas que

estão associados à criação de uma impressão positiva em uma entrevista na mídia.

Embora as dicas não verbais sejam geralmente aceitas como um elemento importante nas entrevistas da mídia e sejam normalmente incluídas no treinamento, a precisão dos treinadores na detecção desses sinais é incerta, pois a inferência de emoções é subjetiva por natureza (Vrij et al., 2000).

### Detecção automatizada de sinais não-verbais Os

desenvolvimentos tecnológicos nos campos da computação afetiva e do processamento de sinais sociais (SSP) nos últimos anos permitiram a detecção automática de uma variedade de sinais não-verbais. Vinciarelli et al. (2012) fornecem uma pesquisa aprofundada do SSP e uma revisão dos sistemas de detecção de afeto pode ser encontrada em D'mello e Kory (2015). SSP é uma área de pesquisa que modela a interação humano-humano para desenvolver máquinas emocionalmente inteligentes.

Da perspectiva do SSP, Pentland propôs que essa interação de comportamento vocal, troca de turnos, movimento e postura nas interações sociais representa um segundo canal de comunicação que ele cunhou com o termo sinais honestos (Pentland e Heibeck, 2010). Esses sinais, que ele identifica como mimetismo (espelhamento – efeito camaleão), influência, atividade e consistência, são propostos como preditores evolutivamente importantes das características e intenções do parceiro de comunicação (Bilakhia et al., 2015). Sung e Pentland (2005) e Curhan e Pentland (2007) fornecem uma série de exemplos empíricos em que sinais honestos predizem resultados de tarefas de comunicação.

Vários estudos usaram essa tecnologia para investigar a comunicação não-verbal em uma variedade de interações. Isso inclui ambientes médicos (Hart et al., 2016), entrevistas de emprego (Frauendorfer et al., 2014; Naim et al., 2016), ensino (Chen et al., 2011, 2015; Bahreini et al., 2016), e melhorar a comunicação social em indivíduos com autismo (Bernardini et al., 2014; Chen et al., 2016). No entanto, não temos conhecimento de pesquisas anteriores com tecnologia automatizada considerando especificamente os sinais sociais necessários para uma comunicação eficaz em uma entrevista na mídia.

Normalmente, a pesquisa no domínio SSP se concentra em canais únicos ou sinais de um único indivíduo, em vez de sinais recíprocos (Kim e Suzuki, 2014). É importante observar a troca recíproca de sinais entre um emissor e um receptor, pois esse intercâmbio influencia o comportamento. Além disso, a pesquisa geralmente se baseia em indicações gerais da literatura sobre quais sinais representam um bom desempenho na tarefa de comunicação, em vez de definir 'bom' em relação ao conjunto específico de sinais detectados pela tecnologia.

Rasipuram e Jayagopi (2018) prevêem o desempenho da comunicação capturando canais multimodais durante uma entrevista face-a-face e uma entrevista de interface. Os sinais capturados incluíam movimentos, expressão facial, gestos com as mãos, postura, contato visual, recursos verbais e atenção. Os pesquisadores descobriram que os participantes tinham uma taxa de fala ideal e se comunicavam melhor na entrevista face a face do que na entrevista de interface.

Essa descoberta sugere que a comunicação é melhor quando múltiplos sinais podem ser vistos em uma interação por ambos os interlocutores. Isso é consistente com a teoria de Adams e Kveragas de que a integração visual de combinações de pistas sociais é necessária para a adaptação comportamental em resposta aos outros (Adams e Kveraga, 2015).

Mais recentemente, a tecnologia comercial de prateleira foi desenvolvida e está disponível para todos os usuários e permite que os usuários produzam dados localmente usando o classificador disponibilizado (Dupré et al., 2018). Isso permite que os usuários acessem facilmente seus dados. Alguns exemplos dessas tecnologias são Emotients FACET ou Affectiva que capturam expressões faciais (Stöckli et al., 2018), Microsoft Kinect para capturar movimentos corporais (Barmaki, 2016), Sociometric Badges para medir interações entre duas ou mais pessoas (Zhang et al., 2018a) e obtenção do movimento das mãos usando acelerômetros (Koskimäki et al., 2017).

A detecção bem-sucedida de sinais sociais associados a um bom desempenho em uma entrevista na mídia usando essa tecnologia pode ter muitas aplicações potenciais. Em primeiro lugar, tem potencial para apoiar um treinador humano, uma vez que os treinadores podem não ser capazes de observar e considerar todas as dicas que podem afetar a comunicação eficaz e o desempenho individual atualmente depende muito da experiência do treinador (Aspegren, 1999). Em segundo lugar, pode selecionar objetivamente os sinais sociais necessários para uma comunicação eficaz em várias interações sociais (Naim et al., 2016).

### A pesquisa atual Em suma, os

contextos em que os sinais sociais foram investigados são entrevistas de emprego, falar em público e na sala de aula (Bahreini et al., 2017). Pesquisas anteriores limitam-se à análise unimodal de interações sociais, mas pesquisas mais recentes fornecem evidências de que uma abordagem multimodal é mais eficaz para sintetizar e interpretar interações sociais. Há pouca ou nenhuma pesquisa investigando os sinais sociais apropriados para uma comunicação eficaz no treinamento de habilidades de mídia, o que é importante devido à natureza da comunicação nesse cenário.

O objetivo deste artigo é duplo:

- (1) Investigue quais combinações de sinais são relevantes em uma entrevista na mídia.
- (2) Apresentar um possível método mais objetivo de capturar sinais sociais durante entrevistas na mídia, em oposição aos métodos tradicionais de assistir a um vídeo.

Conduzimos um estudo para investigar os sinais associados ao bom desempenho em entrevistas de habilidades de mídia, detectando automaticamente uma variedade de sinais sociais (incluindo comportamento recíproco em relação ao entrevistador) durante o contexto de exercícios de treinamento de mídia e analisamos como eles previam um bom e mau desempenho conforme julgado por avaliadores humanos.

## MATERIAIS E MÉTODOS

A pesquisa atual aplicou a detecção automática de sinais sociais em uma **entrevista face a face na câmera**, esta seção



detalha o desenho do estudo (consulte a seção “Desenho do estudo”), as características dos participantes (consulte a seção “Participantes”), a tecnologia usada para capturar sinais sociais durante as entrevistas (consulte a seção “Tecnologia de detecção de sinais não verbais pronta para uso”), como o desempenho foi avaliado por avaliadores humanos (consulte a seção “Medidas subjetivas de habilidades de comunicação”) e como os dados foram coletados descrevendo o procedimento junto com o layout do estudo (consulte a seção “Detalhes do workshop de habilidades de procedimento e mídia”).

Desenho do estudo A

pesquisa atual explorou uma interação diádica durante um ambiente de entrevista na mídia, onde os participantes foram entrevistados por um jornalista em entrevistas face-a-face na câmera. Os sinais que foram detectados automaticamente foram expressões faciais, sinais vocais, 'sinais honestos' e gestos com as mãos. O desempenho da comunicação durante as entrevistas foi julgado por avaliadores humanos. Posteriormente, usando essas classificações, as entrevistas foram categorizadas em comunicadores eficazes e fracos. Os dados foram então explorados para identificar as relações entre os sinais capturados e os julgamentos humanos de desempenho. Os dados foram mais explorados para identificar as relações entre os sinais detectados e os julgamentos humanos. Os detalhes destes podem ser vistos nas seções a seguir.

Participantes Um

total de 39 participantes foi recrutado para participar de um treinamento de entrevista na mídia em uma universidade de Londres (17 homens e 22 mulheres; a idade variou de 18 a 56). Todos os participantes eram estudantes de pesquisa ou funcionários da pesquisa e nenhum deles tinha deficiência social. Foram realizadas duas oficinas, a primeira com 17 participantes (11 homens e 6 mulheres; idade variando de 18 a 65 anos) que incluiu nove participantes falantes nativos de inglês (participantes que declararam ter o inglês como primeira língua) e 10 participantes que não eram falantes nativos de inglês (participantes que declararam que o inglês não era sua primeira língua). A experiência em falar em público variou de 'nenhuma' a 'extensa' e a experiência em entrevistas na mídia variou de 'nenhuma' a 'alguma'. Os papéis que os participantes tiveram dentro da universidade no primeiro workshop incluíram equipe de pesquisa (5), estudante de pesquisa (10), equipe profissional (1) e aluno de ensino (1).

A segunda oficina contou com 22 participantes (6 homens e 16 mulheres; idade variando de 18 a 55 anos), incluindo 6 falantes nativos de inglês e 16 falantes não nativos de inglês.

A experiência em falar em público variou de nenhuma experiência a extensa e a experiência em entrevistas na mídia variou de nenhuma a algumas experiências. Os papéis que os participantes tiveram dentro da universidade no segundo workshop incluíram alunos (3), equipe de pesquisa (1) e alunos de pesquisa (18).

Tecnologia de Detecção de Sinais Não Verbais de

Prateleira Os sinais não verbais que foram detectados durante as entrevistas incluíram sinais vocais, sinais honestos, expressões faciais e movimentos das mãos. Esta seção apresenta a tecnologia comercial usada para capturar esses sinais. As precisões serão relatadas usando medidas de Receptor Operacional

Características (ROC). A medida ROC demonstra a capacidade de diagnóstico de um sistema baseado em uma curva criada pela taxa de verdadeiros positivos contra a taxa de falsos positivos. Quanto mais próxima a pontuação ROC estiver de 1, mais preciso será o classificador, sugerindo que a tecnologia mede o que sugere que ela mede (Macmillan e Creelman, 2004).

Detecção de Comportamento Vocal A

tecnologia QA5 da Nemesysco Ltd foi usada para detectar sinais vocais dos participantes durante as entrevistas. Este software usa algoritmos proprietários de processamento de sinal para extrair parâmetros da voz e classificá-los de acordo com uma variedade de sinais vocais<sup>1</sup>. A Tabela 1 resume as emoções que a tecnologia pretende classificar com uma breve descrição de cada uma.

O escore da área sob a curva ROC para o Nemesysco varia de 0,53 a 0,71 (Lacerda, 2009). No entanto, este estudo não esclareceu qual versão do Nemesysco foi medida. No entanto, alguns sinais captados pelo QA5 foram validados, como 'embaraço' (Han e Nunes, 2010), estresse e excitação (Konopka et al., 2010 conforme citado em Mayew e Venkatachalam, 2010). Foram realizadas pesquisas que usaram o QA5 no desenvolvimento de um robô conversacional (Usui et al., 2008; Hashimoto et al., 2009). O guia de uso do QA5 afirma que o ruído e o ambiente podem influenciar os resultados. Neste estudo, isso foi controlado garantindo-se um ambiente silencioso durante a entrevista.

Para validar os sinais utilizados neste estudo, foi utilizado um software de código aberto para correlacionar os sinais vocais captados pelo QA5 com recursos prosódicos extraídos do Praat. Praat with é um software de extração de voz que pode ser usado para analisar,

<sup>1</sup>Nemesysco.com

TABELA 1 | Definições de rótulos de emoção produzidos por Nemesysco/análise de voz em camadas.

Emoção	Descrição
Energia	Indica se o orador está triste, cansado, entediado, confortável ou com muita energia.
Contente	Indica o quão satisfeita ou feliz uma pessoa está
Chateado	Indica o quão descontente ou triste uma pessoa está
Bravo	Indica o quão zangada uma pessoa está
Estressado	Indica o quão nervosa uma pessoa está
Embaraço	Indica o quão desconfortável uma pessoa está
Pensamento intensivo	Indica a intensidade do pensamento enquanto fala
Atividade de imaginação	Indica se a pessoa está recordando informações ou visualizando algo
Hesitação	Indica o quão confortável uma pessoa está ao fazer a declaração
Incerteza	Indica o quão certa ou incerta uma pessoa é
Excitação	Indica o quão positiva ou negativamente excitada uma pessoa está
Concentração	Indica o quão concentrada a pessoa está
Excitação	Indica profundo e profundo interesse na conversa
emoção extrema	Indica atividade emocional geral
Atividade cognitiva	Atividade cognitiva geral
razão EmCoG	Indica racionalidade

sintetizar e manipular a fala (Boersma e Van Heuven, 2001). Uma análise de correlação foi realizada para validar as características coletadas pela Nemesysco Ltd. As características vocais extraídas do Praat foram altura (média e máxima), intensidade (média, energia, mínima e máxima). O pitch é definido como a velocidade de abertura e fechamento das pregas vocais, também conhecido como frequência fundamental (Giles et al., 1979). Frequência fundamental e intensidade são conhecidas por serem variáveis importantes na comunicação de emoções na fala (Ramdinmawii et al., 2017).

O valor médio do tom para alto-falantes masculinos é normalmente de 100 a 180 Hz e, para mulheres, de 160 a 300 Hz.

Um pitch médio alto tem sido associado a estresse e excitação (Sondhi et al., 2015). A intensidade está associada ao volume da voz e está associada a uma variedade de emoções, incluindo estresse psicológico (Van Lierde et al., 2009).

A Tabela 2 mostra que 'estressado', 'chateado', 'pensamento intensivo', 'imaginação', 'energia', 'animado', 'relação emo cog', 'concentração' e 'emoção extrema' é consistente com características prosódicas extraídas em Praat que são consistentes com a literatura descrita. A Tabela 2 mostra os resultados da correlação.

Para gravar a análise de voz durante as entrevistas, um gravador portátil Zoom H4N Pro foi usado para gravar os sinais de voz. A voz do entrevistador foi editada usando o software Audacity versão 2.1.1 antes do pós-processamento da voz do participante usando o QA5 da Nemesysco Ltd.

Deteção de Sinal Honesta

Pentland e Heibeck (2010) propõem que existem quatro sinais honestos que estão presentes em todas as interações sociais e revelam as atitudes inconscientes de uma pessoa; (1) mimetismo, (2) consistência, (3)

atividade e (4) influência. Os crachás sociométricos foram desenvolvidos por Pentland para detectar uma variedade de sinais hipotéticos por Pentland relacionados a 'sinais honestos'. (ver Pentland e Heibeck, 2010 para uma discussão mais aprofundada). Badges sociométricos têm sido usados para detectar sinais em interações diádicas (Paxton et al., 2015; Zhang et al., 2018b; Holding et al., 2019). A pontuação ROC para esses emblemas foi relatada em 0,8 (Zhang et al., 2018b).

Sinais honestos são detectados por quatro sensores: um microfone, um sensor infravermelho, um detector Bluetooth e um detector de movimento (Olguin e Pentland, 2007). O microfone detecta tons vocais e não conteúdo (Tabela 3, Recursos L – U). O sensor infravermelho capta o movimento relativo a outros interlocutores (Tabela 3, Características E, F, J, K). O sensor Bluetooth detecta outros usuários de crachá. Cada crachá tem aproximadamente o tamanho de um crachá de identidade e é usado ao redor do pescoço. A Tabela 3 lista os sinais que podem ser extraídos dos dados sociométricos do crachá.

Crachás sociométricos foram usados tanto pelos participantes quanto pelos entrevistadores durante as entrevistas. Após a entrevista, os dados armazenados localmente nos crachás foram exportados como reuniões estruturadas (pois os participantes se enfrentaram em uma única reunião)

TABELA 3 | Definições de sinais produzidos por emblemas sociométricos.

Característica	Descrição
A) Movimento corporal	Magnitude de aceleração normalizada em 3 eixos de movimento
B) Atividade de movimento corporal	Valor absoluto da primeira derivada do acelerômetros energia
C) Taxa de movimento do corpo	Indica a direção da mudança no nível de atividade (em comparação com a primeira derivada)
D) Consistência do movimento corporal	Consistência do movimento ao longo da interação
E) Espelhamento do movimento corporal	Imitando o movimento do corpo de outros usuários de crachás
F) Atraso do espelho de movimento do corpo	Atraso na imitação do movimento do corpo
G) Postura de frente para trás	Orientação do painel frontal traseiro
H) Atividade postural	Velocidade angular absoluta
I) Taxa de postura	Aceleração angular
J) Espelhamento de postura	Imitando a postura de outros usuários de crachás
K) Atraso do espelho de postura	Demora na imitação da postura
L) Interrupções bem-sucedidas	Número de interrupções bem-sucedidas feitas pelo usuário dos crachás
M) Interrupções sem sucesso	Número de interrupções malsucedidas feitas pelo usuário do crachá
N) Velocidade de tomada de curva	Indica a velocidade de tomada de turno em uma conversa
O) Sobreposição	Quantidade total de fala enquanto outra pessoa também está falando
P) Fala total	Quantidade total de fala combinada (fala e sobreposição combinadas)
Q) Frente de volume	Valor absoluto médio da amplitude do microfone frontal
R) Frente de consistência de volume	Medição da mudança no volume da fala
S) Passo frontal	Tom da voz do microfone frontal correlacionada com a frequência fundamental do sinal de voz
T) Espelhamento de volume	Imitando o volume de outros usuários de crachás
U) Atraso do espelhamento de volume	Atraso na imitação do volume de outros usuários de crachá

TABELA 2 | Resultados de correlação entre Nemesysco Ltd e um software de código aberto comumente usado.

Resultados da correlação

Característica	Sub-recurso	correlação LVA
Intensidade	Significa	Estresse (r = 0,506, p = 0,002) Perturbação (r = 0,602, p < 0,001)
	Energia	Estressado (r = 0,502, p = 0,002)
		Chateado (r = 0,520, p = 0,002)
	Mínimo	Estressado (r = 0,411, p = 0,016)
		Pensamento intensivo (r = 0,352, p = 0,041) Imaginação (r = 0,501, p = 0,003)
		Energia (r = -0,348, p = 0,044) Animado (r = -0,514, p = 0,002) EmoCogRatio (r = -0,388, p = 0,023)
Fundamental frequência	Máximo	Estressado (r = 0,435, p = 0,010)
		Chateado (r = 0,499, p = 0,003)
		Imaginação (r = 0,379, p = 0,028)
	Significa	Estressado (r = 0,534, p = 0,001)
		Energia (r = 0,742, p < 0,001)
		Excitação (r = 0,471, p = 0,005)
		Concentração (r = 0,519, p = 0,002) EmoCogRatio (r = 0,641, p < 0,001) Pensamento intensivo (r = -0,622, p < 0,001)
		Imaginação (r = -0,591, p < 0,001)
	Máximo	Pensamento intensivo (r = -0,369, p = 0,032)

com uma resolução de intervalos de 1 s (Sociometric Solutions, 2015).

Os crachás usados pelo treinador e pelo participante foram sincronizados usando o software Sociometric Solutions (Sociometric DataLab Enterprise Edition 3.1.2824).

#### Detecção de expressão facial As expressões

faciais foram detectadas usando o software iMotions Biometric Research Platform 6.4 e analisadas usando o Affdex da Affectiva. Este software comercial usa um Sistema de Codificação de Análise Facial Emocional (EmFACS) que produz 7 expressões faciais (triste, alegria, raiva, medo, nojo, desprezo e surpresa) que os humanos usam para se comunicar (Ekman e Friesen, 1971).

O sulco da testa, o sorriso afetado e o sorriso também foram avaliados, pois são considerados importantes para uma entrevista na mídia (Taylor, 2015). A Affdex pela pontuação ROC da Affectiva foi relatada como 0,8 para alegria, nojo, desprezo e surpresa (Dupré et al., 2018). As entrevistas foram gravadas com uma câmera portátil Sony PJ220. Todas as edições nas gravações foram feitas usando o Adobe Photoshop. As gravações de vídeo foram importadas para o iMotions e pós-processadas usando o Affdex.

#### Detecção de movimentos/gestos das mãos O Shimmer 3 Unit+

foi usado para capturar os movimentos das mãos.

O dispositivo Shimmer contém um acelerômetro de direção de 3 pontos (x, y, z) que foi usado para obter uma estimativa dos movimentos das mãos usados durante as entrevistas cujo uso dos gestos das mãos será inferido.

**Sequência de Eventos e Timestamps** Todas as gravações dos canais de comunicação foram sincronizadas para 1 s timestamp devido às capacidades das diferentes tecnologias.

Algumas tecnologias não foram capazes de registrar timestamps mais curtos.

Os dados foram analisados como se exibições de sinais sociais ocorressem simultaneamente dentro do período de 30 segundos (Paxton et al., 2015; Naim et al., 2016; Zhang et al., 2018a; Holding et al., 2019; Pereira e Hone, 2021).

**Medidas subjetivas de habilidades de comunicação** A avaliação do desempenho de comunicação dos participantes por avaliadores

humanos foi importante, pois isso reduziria o viés ao identificar comunicadores eficazes e fracos. Também foi feita uma avaliação para identificar relações entre os padrões de sinais emocionais/não-verbais e as avaliações de desempenho dos treinandos, avaliadas por humanos. Para obter julgamentos objectivos do desempenho dos formandos, as entrevistas aos participantes foram avaliadas pelo formador e, posteriormente, por três observadores neutros, utilizando um questionário de avaliação da comunicação (ver secção "Escala de avaliação das competências de conversação").

Várias abordagens foram tomadas para reduzir a subjetividade nas avaliações de desempenho. Em primeiro lugar, porque os treinadores interagiram com os treinandos no dia do treinamento, o que provavelmente teria influenciado suas pontuações como resultado de uma impressão de interação que poderia influenciar as avaliações de julgamento (Meissel et al., 2017), avaliações adicionais foram obtidas por três observadores neutros que não estiveram presentes no dia da formação (Naim et al., 2016).

As classificações obtidas de três observadores neutros destinavam-se a atuar como uma audiência, podendo revisar ambas as entrevistas várias vezes para uma classificação mais completa, além de fornecer classificações mais realistas (Naim et al., 2016). Em segundo lugar, para reduzir ainda mais o potencial de viés de classificação, os observadores neutros eram cegos para as classificações fornecidas pelo treinador.

#### Instrutores e Observadores Neutros Os

jornalistas do primeiro workshop eram homens e mulheres com mais de 20 anos de experiência de campo e haviam realizado o primeiro workshop de treinamento em habilidades de mídia. Os jornalistas deste workshop forneceram feedback aos participantes sobre seu desempenho após as entrevistas. As entrevistas foram divididas igualmente entre os dois jornalistas. Tanto o jornalista quanto os observadores neurais puderam reproduzir e pausar suas entrevistas. A jornalista que conduziu as entrevistas no segundo workshop era uma mulher com 4 anos de experiência de campo e havia conduzido todas as entrevistas.

Os observadores neutros recrutados para avaliar o desempenho da comunicação a partir das gravações das câmeras não foram treinados no que é considerado 'comunicação eficaz' e foram tratados como um membro da população em geral. Os três observadores neutros recrutados para o primeiro workshop eram diferentes dos observadores neutros do segundo workshop. Nem os jornalistas nem os observadores neutros sabiam quem havia sido rotulado como um comunicador eficaz ou um comunicador ruim.

#### Escala de Avaliação de Habilidades de

Conversação Classificações humanas subjetivas de habilidades de comunicação foram obtidas usando a Escala de Avaliação de Habilidades de Conversação (CSRS) (Spitzberg e Adams, 2007). O CSRS tem duas seções de classificação: uma escala de 25 itens que classifica as características de comunicação verbal e não verbal e uma escala de 5 itens que mede o desempenho geral da comunicação (classificações molares). Como este estudo incluiu um rádio e uma entrevista face-a-face, as pontuações gerais de desempenho de comunicação (classificações molares) foram usadas, pois isso não inclui nenhum item da escala que incluía medidas interpessoais de comunicação que não seriam visíveis para observadores neutros quando ouvir a entrevista de rádio e, portanto, não pode ser avaliado. Os avaliadores foram solicitados a se concentrar em dicas não verbais enquanto assistiam aos vídeos.

O CSRS é uma medida de habilidades interpessoais e é considerado aplicável em 'praticamente todas as interações de conversação face a face' (Spitzberg e Adams, 2007). Evidências de sua confiabilidade e validade foram encontradas em vários ambientes, incluindo ambientes educacionais, entrevistas de emprego e conversas para conhecê-lo (Spitzberg e Adams, 2007). Embora não tenhamos encontrado exemplos específicos de seu uso na avaliação de habilidades de mídia, também não fomos capazes de identificar quaisquer outras ferramentas validadas e consideradas relevantes para este contexto. A confiabilidade interna do CSRS tem estado consistentemente acima de 0,85 e frequentemente acima de 0,90. A confiabilidade entre avaliadores foi avaliada e encontrou confiabilidades aceitáveis acima de 0,75 (Spitzberg e Adams, 2007). As classificações molares foram preenchidas pelos treinadores e três observadores neutros para avaliar o desempenho das habilidades de comunicação nas entrevistas na câmera.

## Procedimento e detalhes do workshop de habilidades de mídia

O estudo ocorreu no campus de uma universidade de Londres no contexto de dois dias de treinamento de mídia para pesquisadores.

Os três primeiros workshops foram conduzidos por profissionais de media training com mais de 20 anos de experiência profissional em jornalismo. Um total de três dias de treinamento ocorreu no período de abril de 2017 a junho de 2017, com o número de participantes por dia variando de 5 a 6. Toda a coleta de dados ocorreu em uma sala de seminário padrão da universidade com mesas, cadeiras e um projetor. Os três workshops restantes foram conduzidos por um jornalista em início de carreira com 6 anos de experiência na área.

Estes workshops decorreram no período de novembro de 2017 a dezembro de 2017. A partir deste momento, os jornalistas serão referidos como formadores.

Antes de participar do treinamento, os participantes foram solicitados a fornecer um breve resumo de sua pesquisa que seja compreensível para uma população não especializada, incluindo a importância e a pior pergunta esperada em uma entrevista na mídia. Isso foi feito para ajudar os treinadores a se prepararem para a realização de entrevistas práticas de mídia adaptadas aos perfis de pesquisa e trabalho de cada participante.

Na chegada ao dia do treinamento, os participantes foram totalmente informados sobre o estudo e o consentimento formal foi coletado, juntamente com informações demográficas (função no trabalho, gênero, idade e etnia, presença de deficiência social/de comunicação e experiência anterior de apresentação). Caso os participantes não desejassem dar consentimento para a gravação de sinais sociais, eles tiveram a opção de participar com os sistemas desligados durante as entrevistas sem penalidade. Todos os participantes deram consentimento para gravar os sinais.

Após uma introdução, os participantes participaram de uma introdução de estilo de palestra de 45 minutos a 1 hora para uma entrevista de mídia eficaz

habilidades de comunicação. A palestra foi apresentada em grupo. Os participantes receberam intervalos de tempo individuais durante o dia para voltar a fazer entrevistas práticas com os treinadores.

As entrevistas práticas foram realizadas individualmente, e duas entrevistas práticas foram realizadas para cada participante. A primeira foi simular uma entrevista de rádio, de modo que os participantes se sentassem frente a frente com um gravador de voz sobre a mesa. Nenhuma câmera foi ligada

durante as entrevistas para evitar qualquer influência que isso possa ter no desempenho. A segunda entrevista prática foi uma simulação de uma entrevista na câmera, então a câmera foi posicionada atrás do jornalista e ao lado do participante. Os participantes foram informados de que a câmera foi colocada atrás do jornalista que estava gravando como se fosse uma transmissão. A configuração da sala é ilustrada na **Figura 1**.

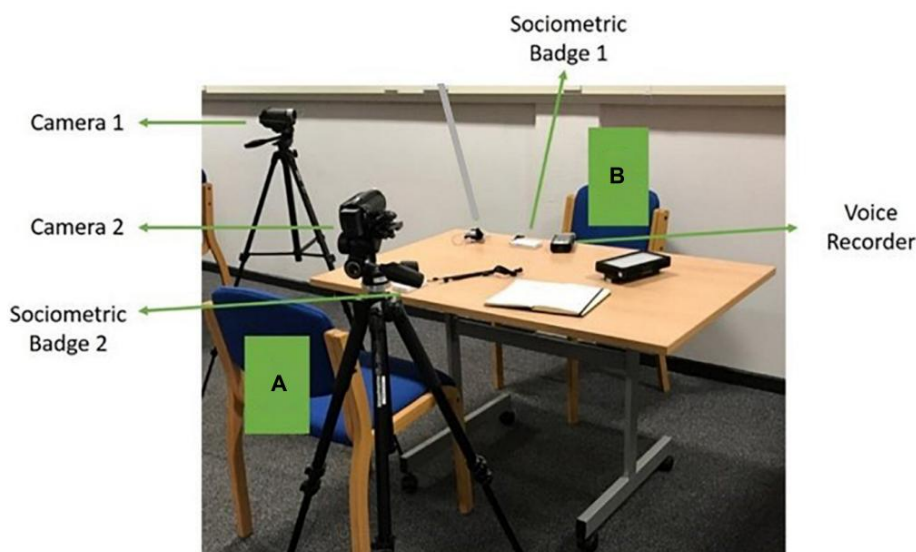
Um dos treinadores atuou como entrevistador para o propósito das entrevistas práticas com a mídia.

Antes do início das entrevistas, os participantes foram conectados a um dispositivo Shimmer 3 GSR e tanto o participante quanto o entrevistador colocaram um crachá sociométrico. A sala também foi montada com mais equipamentos de gravação para permitir a detecção de sinais/emoções sociais, conforme mostrado na **Figura 1**.

Durante as entrevistas práticas, foram feitas aos participantes perguntas individualmente relevantes sobre sua pesquisa. A primeira pergunta pedia aos participantes que explicassem sua pesquisa. Essas perguntas foram baseadas no material fornecido pelos participantes.

A dificuldade da questão aumentou à medida que as entrevistas avançavam. Cada entrevista durou entre 5 e 8 minutos.

As gravações das entrevistas foram reproduzidas para os participantes após cada entrevista e eles receberam feedback de desempenho dos treinadores que puderam reproduzir as entrevistas, o que permitiu que o treinador e o aluno pausassem e rebobinassem o vídeo para um feedback de desempenho eficaz. Os treinadores foram então solicitados a preencher o CSRS, que é uma medida padronizada de Habilidade em comunicação.



**FIGURA 1** | Estude o layout das sessões de treinamento de rádio e na câmera. As câmeras foram ligadas apenas para gravação durante a entrevista na câmera. (A) Cargo de jornalista. Ambas as câmeras estão voltadas para o participante (B) para um reconhecimento facial post hoc mais preciso. O gravador de voz adicionado foi para uma gravação de entrevistas de melhor qualidade.



Após a conclusão do estudo, os participantes receberam uma breve declaração de encerramento, lembrando-os do objetivo da pesquisa. Os participantes foram reembolsados em £ 5/h pelo reconhecimento de seu tempo.

A natureza subjetiva do julgamento humano torna difícil estabelecer a verdade básica para as entrevistas. O formador interagiu com os formadores durante a palestra, durante a entrevista e forneceu feedback após cada sessão. Essa quantidade de interação pode ter influenciado as avaliações dos treinadores. Portanto, a fim de remover possíveis vieses, as entrevistas gravadas foram posteriormente avaliadas por três observadores neutros também usando o CSRS. Outros benefícios disso são que os observadores neutros podem revisar o material várias vezes, o que lhes permite avaliar as entrevistas de forma mais completa. As gravações de vídeo das entrevistas na câmera foram apresentadas a observadores neutros para obter julgamentos do desempenho da comunicação. As avaliações desses observadores provavelmente seriam semelhantes às avaliações do público de uma entrevista na mídia, em oposição às avaliações de especialistas (Naim et al., 2016). Os observadores neutros puderam interagir com os vídeos pausando, rebobinando e encaminhando os vídeos de cada participante.

Cada observador neutro trabalhou individualmente e estava cego para as avaliações fornecidas pelos outros.

Esta pesquisa foi conduzida de acordo com a Declaração de Helsinque e a aprovação ética foi obtida do Comitê de Ética em Pesquisa do Ministério da Defesa, bem como do Comitê de Ética em Pesquisa de Universidades.

## RESULTADOS

**Avaliações subjetivas de habilidades de comunicação** Entrevista na câmera A comunicação dos participantes foi avaliada usando o CSRS pelos treinadores e, posteriormente, por três observadores neutros. **A consistência interna foi calculada por meio do Alfa de Cronbach.** Em seguida, uma média composta da classificação geral das habilidades de comunicação (com base nas cinco classificações molares) foi obtida para as classificações do treinador e três observadores neutros. A confiabilidade entre avaliadores foi conduzida para calcular a concordância entre os avaliadores usando a correlação intraclasse com uma abordagem mista bidirecional (Mandrek, 2011). A consistência interna foi alta para avaliações de comunicação para todos os avaliadores de comunicação (escores molares,  $n = 5$ ) foi  $\bar{y} = 0,95$ . A correlação intraclasse foi de 0,78 com intervalo de confiança de 95% de 0,603 a 0,870 [ $F(4,289)$ ,  $p < 0,001$ ]. Este acordo moderado garante uma média ponderada (Mandrek, 2011). A mediana do conjunto de dados foi de 24,33, que **estabeleceu comunicadores eficazes ( $M = 28,35$ ;  $SD = 3,22$ ) e fracos ( $M = 19,60$ ;  $SD = 3,10$ ).**

### Exibições de sinais sociais durante a

#### comunicação

Dados ausentes Casos com dados ausentes de qualquer canal foram excluídos da análise. Um total de seis participantes foi excluído (três devido a gravações de vídeo de baixa qualidade para expressões faciais, dois devido à falta de dados de gestos manuais e um devido à falta de dados sociométricos).

dados do crachá). Isso resultou em um tamanho de amostra de 33 participantes incluídos na análise.

### Pré-processamento de dados – Normalização

Os dados do sinal social foram normalizados usando os valores mínimo e máximo dos conjuntos de dados, resultando em um intervalo de conjuntos de dados de 0 a 1 (Gao et al., 2012). A fórmula pode ser vista abaixo.

$$x = \frac{x - \min(x)}{\max(x) - \min(x)}$$

### Thin Slices of Behavior Research

descobriu que os primeiros 30 s de uma interação foram mais eficazes na avaliação de julgamentos e percepções sobre as pessoas, pois os avaliadores de desempenho baseiam suas pontuações nos estágios iniciais de uma interação (Sullivan, 2018). As impressões são normalmente feitas durante essa escala de tempo, mesmo que uma interação completa possa ocorrer. Isso sugere que a resposta dos entrevistados à primeira pergunta da entrevista pode ter influenciado os observadores na formação de julgamentos iniciais sobre suas habilidades de comunicação. É por essas razões que decidimos investigar os primeiros 30 s das entrevistas gravadas. Além disso, em um contexto de entrevista, descobriu-se que os primeiros 30 segundos são cruciais para a tomada de decisão sobre o candidato, pois o relacionamento é construído nos primeiros 30 segundos (Forbes e Jackson, 1980; Duggan e Parrott, 2001).

Os primeiros 30 segundos de uma entrevista na mídia são benéficos para estabelecer padrões nos sinais sociais associados ao julgamento do desempenho da entrevista na mídia. Os primeiros 30 s nas entrevistas foram suficientes para obter a primeira pergunta e resposta em cada entrevista. Conforme observado anteriormente, a pesquisa mostrou que a primeira parte da entrevista permite que os entrevistadores façam um julgamento/formem uma impressão do entrevistado (Sullivan, 2018). O mesmo pode ser dito para entrevistas na mídia (Taylor, 2015), falar em público (Chollet et al., 2015), como nosso comportamento de fala prediz como somos vistos em mídias sociais online (Park et al., 2016) e em entrevistas de emprego (Nguyen e Gatica-Perez, 2015; Naim et al., 2016). Além disso, uma meta-análise descobriu que as classificações de previsão não diferem entre 30 segundos da entrevista e 5 minutos (Ambady e Rosenthal, 1992).

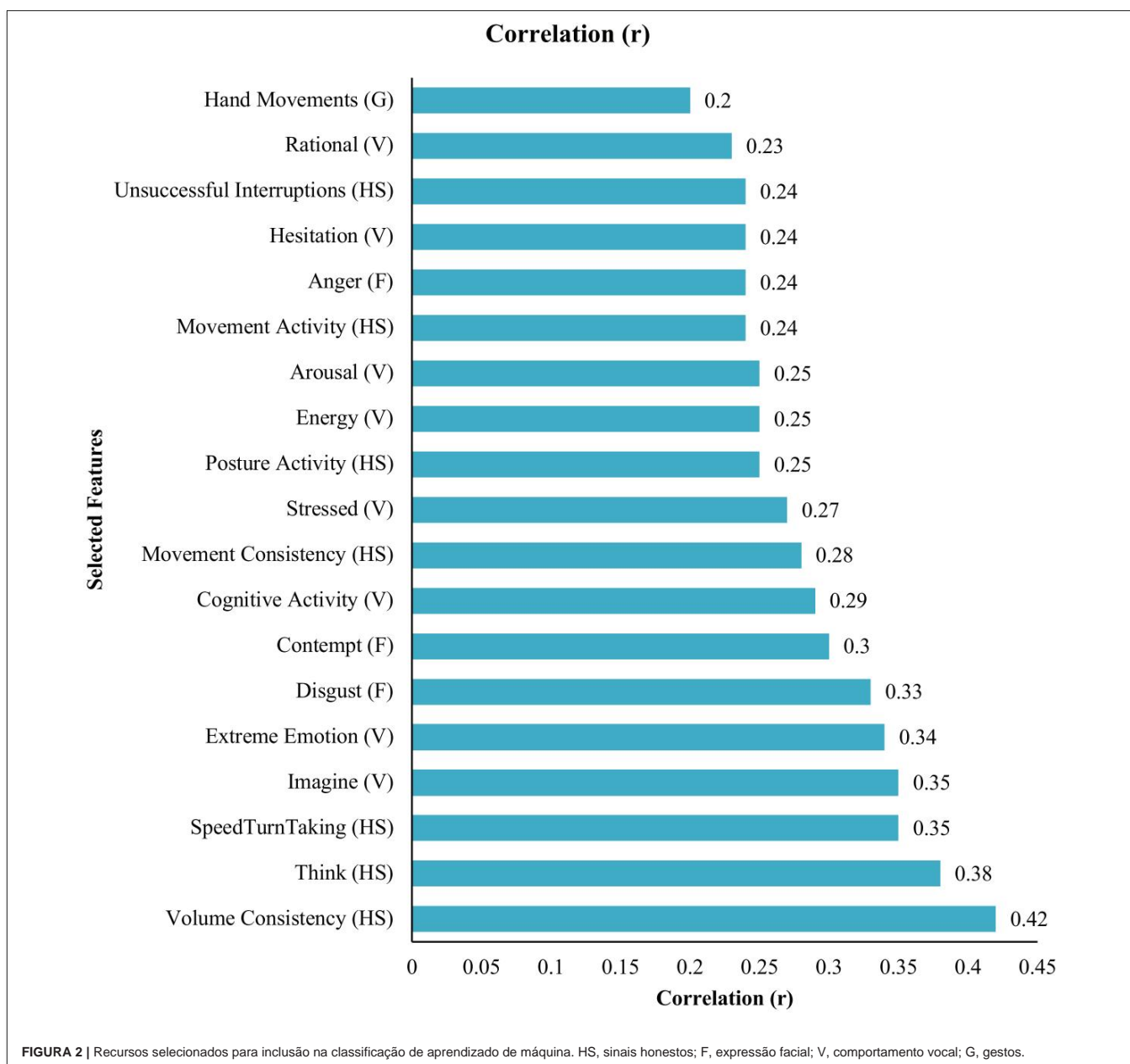
### Técnicas de classificação de aprendizado de máquina

**Estabelecendo uma verdade básica** Uma média das classificações dos observadores neutros (juizes) foi obtida para cada participante (Naim et al., 2016). As avaliações dos observadores neutros foram coletadas à medida que eram tratadas como uma audiência.

Uma mediana para o conjunto de dados foi identificada para que cada entrevista pudesse ser rotulada como comunicadores eficazes ou ruins, formando uma verdade básica para as técnicas de aprendizado de máquina que serão usadas. Isso foi feito para estabelecer um ponto médio no conjunto de dados para estabelecer classificações altas e baixas de comunicação. Um valor alto indica comunicação eficaz e um valor baixo indica comunicação ruim. O limite entre bom e ruim na entrevista para a câmera foi 24,33. As classificações de rádio não foram incluídas.

### Seleção de recursos

As relações entre padrões de sinais não-verbais e avaliações de desempenho de trainees foram exploradas usando Weka



GUI Versão 3.8. Os recursos foram selecionados usando uma seleção de recursos baseada em correlação (CFS). Qualquer coisa abaixo de um ponto de corte de 0,2 foi excluída. Este método seleciona os recursos que são altamente correlacionados com os dados rotulados e não correlacionados uns com os outros (Witten e Frank, 2002). O CSF foi aplicado a todos os canais de comunicação simultaneamente. O método CSF foi usado para recursos selecionados para inclusão na análise de aprendizado de máquina em que uma classificação binária de classificações de comunicação boa e ruim. As características selecionadas com base nos métodos CSF podem ser vistas na **Figura 2**.

#### Classificação de aprendizado de máquina

Usando os dados coletados e pré-processados, o desempenho foi avaliado usando os seguintes classificadores (usados com padrão

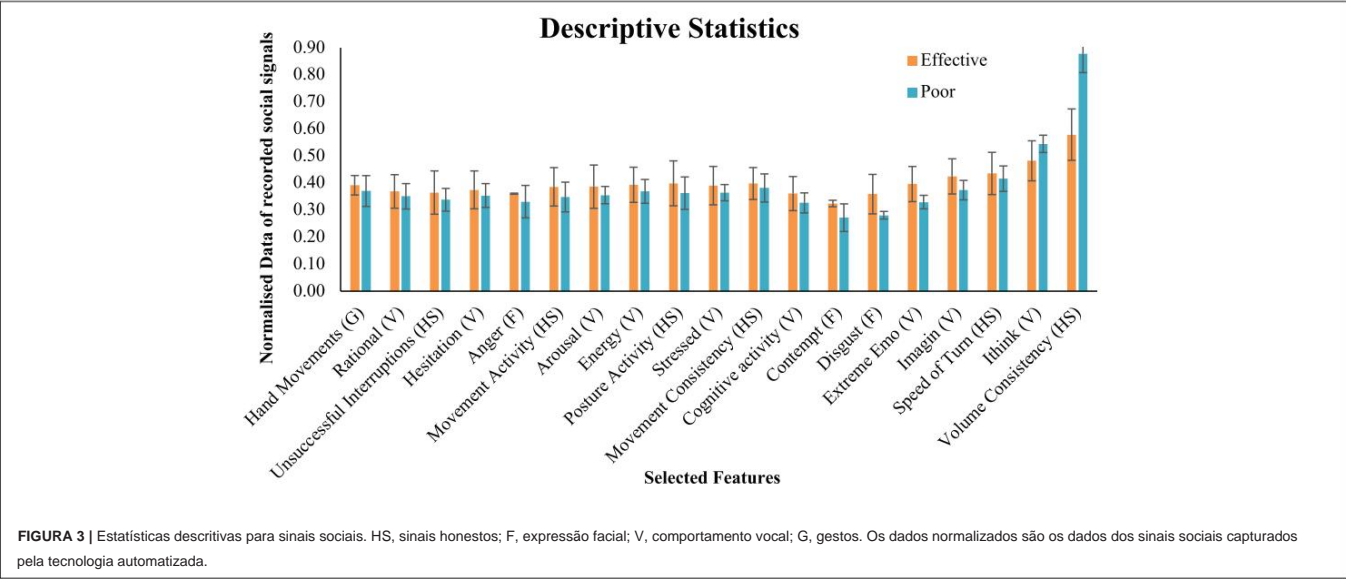
parâmetros salvo indicação em contrário) Logistic Regression, Naïve Bayes, Decision Tree, k-Nearest Neighbor com um parâmetro onde  $k = 3$  e Support Vector Machine (Poly Kernel). O número de participantes classificados como comunicadores eficazes foi (15) e o número de participantes classificados como comunicadores ruins foi (18).

A validação cruzada de deixar um de fora foi usada para dados não balanceados, bem como dados com um tamanho de amostra pequeno (Witten e Frank, 2002). A validação cruzada leave-one-out é onde os algoritmos são aplicados uma vez para cada instância, usando todas as outras instâncias como um conjunto de treinamento e usando o selecionado como um conjunto de teste de item único (Witten e Frank, 2002). A medida F (também conhecida como pontuação F1 ou pontuação F) foi selecionada como o desem

TABELA 4 | Resultados da classificação de aprendizado de máquina.

Algoritmo de classificação de aprendizado de máquina	Precisão	F-Measure (média ponderada)	ROC (média ponderada)
regressão logística	61%	0,60	0,59
<b>Baías ingênuas</b>	<b>78%</b>	<b>0,76</b>	<b>0,79</b>
árvore de decisão	55%	0,51	0,50
k-vizinhos mais próximos (k = 3)	67%	0,64	0,66
Máquina de vetores de suporte (PolyKernel)	64%	0,63	0,62

Negrito destaca o melhor resultado.



métrica de avaliação, pois é adequada para dados de classificação desequilibrados e combina precisão e revocação (Goutte e Gaussier, 2005). A análise foi feita usando Weka 3.8.4. De acordo com a **Tabela 4**, o melhor resultado para o conjunto de dados atual é o Naïve Bayes, que produziu uma pontuação de precisão de 78%, uma medida F de 0,76.

Exibição de Sinal Social – Diferenças entre Grupos

Estatísticas descritivas para comunicadores eficazes e ruins para cada sinal social podem ser vistas na **Figura 3**. As barras de erro exibidas na tabela são erros padrão.

Uma análise estatística mais formal foi realizada para testar se os sinais individuais selecionados diferiam entre desempenho de avaliações de entrevista eficaz e ruim. Um teste U de Mann-Whitney foi usado para avaliar as diferenças entre as exibições de recursos do grupo. Os resultados mostraram que a raiva e a consistência do movimento foram significativamente diferentes entre comunicadores eficazes e ruins. Onde aqueles que se comunicaram de forma mais eficaz demonstraram mais raiva do que aqueles que tiveram um desempenho ruim de acordo com observadores neutros. Aqueles que foram classificados como comunicadores mais eficazes exibiram movimentos mais consistentes do que aqueles que eram maus comunicadores. Os resultados podem ser vistos na **Tabela 5**.

TABELA 5 | Resultados do teste Man-Whitney U.

Sinal	Mann-Whitney U	Sig (valor-p)
Movimento da mão (G)	110	0,366
Atividade cognitiva	100	0,206
Interrupções sem sucesso	100	0,196
Hesitação	101	0,219
Raiva	71	0,021*
Atividade de movimento	95	0,148
Excitação	118	0,539
Energia	89	0,100
atividade de postura	99	0,193
Estressado	104	0,262
Consistência de movimento	59	0,006*
Racionalidade	111	0,386
Desprezo	96	0,159
nojo	115	0,470
emoções extremas	90	0,104
Imaginação	91	0,116
Velocidade de TurnTaking	87	0,084
Eu acho que	90	0,104
Consistência de volume	89	0,096

\*Menos de 0,05.

## DISCUSSÃO

A pesquisa no campo da observação de vídeo para entender a interação social é subjetiva e não se adapta a grandes volumes de dados. A tecnologia automatizada pode se apresentar como uma solução possível, detectando objetivamente sinais não verbais e fazendo isso muito mais rápido do que a codificação manual de dados observacionais. Como resultado desse problema, o objetivo de nossa pesquisa foi explorar quais combinações de sinais sociais são mais promissoras para avaliar automaticamente o desempenho do trainee. Os resultados sugerem que o posicionamento corporal, as expressões faciais, os sinais vocais e os gestos das mãos são relevantes para o contexto das entrevistas na mídia. As combinações desses sinais produziram uma previsão de bons e maus desempenhos com uma precisão de 78% e uma medida F de 0,76. Dois sinais sociais sugeriram que havia uma diferença entre comunicadores eficazes e fracos. Comunicadores eficazes demonstraram mais raiva e movimentos mais consistentes nos primeiros 30 segundos de sua entrevista do que aqueles que foram identificados como maus comunicadores. Os resultados do estudo são apresentados com mais detalhes nas seções a seguir e discutidos à luz de pesquisas anteriores.

### Sinais honestos

**Sinais honestos** incluídos na seleção de recursos para a análise multimodal formal incluíram interrupções malsucedidas, atividade de movimento, atividade de postura, consistência de movimento, velocidade de tomada de curva e consistência de volume. A análise formal dos resultados revelou que a consistência do movimento foi significativamente diferente entre os grupos onde aqueles que foram classificados como comunicadores eficazes.

A literatura anterior descobriu que a consistência no movimento sugere que o comunicador está relaxado, calmo e confiante. Isso é particularmente importante em entrevistas na mídia, pois muita inquietação pode sugerir que o entrevistado está desconfortável (Taylor, 2015). Este é um nível de consistência que poderia ter sido identificado pelos juízes como um método eficaz de comunicação (Hill et al., 1981).

### Comportamento Vocal

O sinal vocal, rotulado por Nemesysco, incluído na análise foi a atividade cognitiva. As estatísticas descritivas sugeriram que os comunicadores eficazes exibiam mais atividade cognitiva. No entanto, esta diferença não foi significativamente diferente. Esse resultado sugere que a consideração geral no comportamento vocal é importante no contexto de entrevistas na mídia nos primeiros 30 segundos. Durante os primeiros 30 s da entrevista capturou a primeira pergunta feita pelo jornalista, o que sugere que os entrevistadores foram atenciosos em sua resposta a essa pergunta inicial. As razões pelas quais isso pode não ser significativamente diferente podem ser resultado da análise de fusão, ou seja, exibições vocais de pensamento em combinação com outro sinal social.

### Expressões faciais

As expressões faciais identificadas neste estudo como preditoras de comunicação eficaz ou ruim no contexto de entrevistas na mídia são raiva e repulsa. Os resultados sugerem que aqueles que foram classificados como comunicadores eficazes por observadores neutros

demonstraram mais raiva e nojo do que aqueles que foram classificados como maus comunicadores.

As AUs envolvidas na expressão facial de raiva são AU4, AU5, AU7 e AU23. As AUs envolvidas no nojo são AU9, AU15 e AU16. AU9 e AU4 estão associados ao rebaixamento da sobrancelha. Esse abaixamento da sobrancelha tem sido associado à concentração. Como os dados foram analisados apenas nos primeiros 30 segundos da entrevista, isso pode sugerir que os participantes estavam ouvindo a primeira pergunta feita pelo jornalista ou estavam concentrados (Ekman, 1997).

### Movimentos/gestos manuais

Os gestos manuais foram incluídos na análise de inclusão de recursos. Isso sugere o que a literatura nos informou, que os gestos auxiliam na comunicação (Goldin-Meadow e Alibali, 2013).

Esses resultados sugerem que os wearables podem ser usados para apoiar os apresentadores de mídia, pois é uma intervenção de baixo custo para capturar o uso do gesto da mão. Curiosamente, Damian et al. (2015) desenvolveram um sistema para fornecer feedback em tempo real durante a fala em público baseado em parte na captura de gestos. Essa escolha de projeto foi motivada pela consideração prática do que funcionaria em um ambiente potencialmente ruidoso e não incluiu pré-testes para o que prevê um desempenho 'bom'. No entanto, nossas descobertas fornecem algum suporte empírico para a abordagem escolhida.

### Combinações de Sinais

A combinação de sinais sociais incluiu movimentos de mão (G), racionalidade (V), interrupções malsucedidas (HS), hesitação (V), raiva (F), atividade de movimento (HS), excitação (V), energia (V), atividade postural (HS), estressado (V), consistência do movimento (HS), atividade cognitiva (V), desprezo (F), repulsa (F), emoção extrema (V), imaginação (V), velocidade de troca de turnos (HS), pensamento (V) e consistência de volume (HS). Os sinais incluídos na análise como resultado da análise de inclusão de recursos incluíram sinais honestos, expressões faciais, gestos com as mãos e comportamento vocal. Isso sugere que a comunicação durante entrevistas na mídia é multimodal, o que tem sido sugerido inúmeras vezes (Pantic et al., 2011; Bekele et al., 2013; Potamianos, 2014; D'mello e Kory, 2015; Esposito et al., 2015; Hunyadi, 2019).

A detecção de expressão facial de desprezo e raiva pode sugerir um falso positivo, pois as pessoas costumam franzir a testa ao ouvir alguém. Além disso, o franzido da sobrancelha é uma AU que compõe o desacato. Isso não significa que eles estão com raiva, mas pode ser um sinal de concentração (Rozin e Cohen, 2003). Isso é consistente com a inclusão do recurso 'atividade cognitiva' no processo de seleção de recursos, sugerindo que os participantes estavam ouvindo (franzindo a testa e desprezando) e respondendo de maneira ponderada.

### Limitações atuais do estudo e recomendações para trabalhos futuros

Este estudo exploratório teve um tamanho de amostra relativamente pequeno de 33 participantes. Um aumento no tamanho da amostra em trabalhos futuros seria útil para testar a confiabilidade dos achados aqui descritos. Além disso, um tamanho amostral maior permitiria investigar as diferenças de gênero e avaliar se há



quaisquer diferenças culturais no desempenho. Os resultados devem, portanto, ser interpretados com cautela.

O estudo analisou apenas uma população composta por pesquisadores em início de carreira dentro de um ambiente universitário. Embora, até certo ponto, essa população possa ser vista como representativa do tipo de função profissional em que os funcionários podem ser chamados a participar de entrevistas na mídia, seria interessante confirmar as descobertas para estagiários em outros tipos de organização. Nenhum dos estagiários era especialista em habilidades de mídia, o que poderia ter restringido o alcance do desempenho. Seria interessante em trabalhos futuros incluir participantes experientes e novatos.

No entanto, os resultados são relevantes para um contexto de formação em que os formandos geralmente ainda não são especialistas. Os participantes também receberam perguntas diferentes uns dos outros, devido ao seu próprio histórico de pesquisa. Embora isso tenha aumentado a validade ecológica do estudo, reduziu o grau de controle do experimentador sobre os estímulos e pode ter levado a diferenças de dificuldade e/ou impacto emocional entre diferentes participantes. No entanto, incluímos os primeiros 30 s da entrevista na análise, o que incluiria a pergunta inicial que era para os participantes descreverem sua pesquisa. O trabalho futuro poderia explorar o uso de conjuntos de perguntas mais padronizados.

Uma limitação relacionada à detecção de gestos manuais foi que a tecnologia dos gestos foi amarrada à mão não dominante, o que pode não ser uma representação verdadeira dos gestos manuais. No entanto, os resultados encontrados neste estudo são consistentes com pesquisas anteriores, sugerindo que o uso de gestos manuais é frequentemente percebido por outras pessoas como comunicadores eficazes. Pesquisas futuras podem explorar o uso da detecção de gestos sem contato para evitar esse problema.

Uma limitação potencial para esta pesquisa é que todo o vídeo foi mostrado para observadores neutros enquanto os primeiros 30 s foram incluídos na análise. No entanto, a pesquisa mostrou que não houve diferenças nas previsões com base em 30 s ou 5 min (Ambady e Rosenthal, 1992). No entanto, pesquisas futuras podem explorar e verificar isso.

Uma limitação final deste artigo é que o conteúdo verbal pode ter influenciado as classificações de “comunicação eficaz” que os observadores neutros deram aos participantes. No entanto, uma vez que os pesquisadores incluíram o comportamento vocal como uma variável, o conteúdo não poderia ter se tornado ininteligível (ou seja, splicing ou filtragem aleatória), pois isso não seria adequado para os objetivos deste artigo. Além disso, os avaliadores foram instruídos a se concentrar em recursos não-verbais ao avaliar a eficácia da comunicação.

Há evidências emergentes do valor agregado da combinação de sinais em várias modalidades para melhorar a precisão da classificação (por exemplo, Pantic et al., 2005; Turk, 2014). A abordagem descrita neste artigo facilita a compreensão do valor e dos insights que podem ser obtidos de cada ferramenta e, como tal, é mais relevante para fornecer feedback aos entrevistados, pois eles precisariam saber em quais sinais de cada ferramenta focar para melhorar o desempenho. O trabalho futuro pode comparar os resultados de diferentes ferramentas e avaliar a adequação de cada uma à configuração atual.

O uso de tecnologia comercial de prateleira tem limitações, pois os algoritmos usados para classificar o comportamento não-verbal não estão abertos ao escrutínio direto dos pesquisadores. No entanto, os resultados encontrados foram consistentes com

literatura que suporta a sua aplicabilidade prática neste domínio que valida os resultados atuais.

Outra possível limitação desta pesquisa é que foram avaliados apenas os primeiros 30 s. Os detalhes sobre o desempenho da comunicação dos trainees poderiam ter melhorado ou piorado ao longo da entrevista, o que não foi incluído nesta análise. No entanto, a pesquisa mostrou que julgamentos ou impressões de desempenho são decididos nos primeiros estágios da observação e o desempenho durante o restante da entrevista é tratado como confirmação dos julgamentos iniciais feitos (DeCoster e Claypool, 2004; Sullivan, 2018). Além disso, os primeiros 30 s também foram usados para controlar as perguntas iniciais da entrevista para controlar as restrições em torno das perguntas e permanecer consistentes para todos os participantes. Pesquisas futuras poderiam investigar toda a interação em vez dos primeiros 30 s.

## CONCLUSÃO E CONTRIBUIÇÕES

Neste artigo, investigamos se sinais sociais podem ser detectados em uma interação diádica usando tecnologia automatizada comercial e se boas entrevistas podem ser distinguidas de entrevistas ruins com base em tais sinais. Os resultados desta pesquisa ilustram que várias tecnologias comerciais são capazes de detectar sinais sociais relevantes para o desempenho em uma entrevista na mídia em que há uma troca recíproca de sinais sociais.

Os resultados desta pesquisa têm aplicação potencial em uma variedade de contextos. Eles podem ser usados para auxiliar os treinadores no treinamento convencional de habilidades de mídia, dando-lhes um mecanismo para fornecer aos alunos um feedback mais objetivo sobre seu desempenho não-verbal para aprimorar suas habilidades de comunicação. Nossos resultados podem ajudar os treinadores a escolher as tecnologias prontas mais úteis para usar para apoiar sua função e destacar os sinais mais relevantes para fornecer feedback. Os resultados sugerem que, para entrevistas na tela, os sinais honestos são os sinais mais prevalentes necessários para o conteúdo da entrevista na mídia, seguidos pela expressão facial. A tecnologia utilizada para capturar os gestos das mãos oferece uma boa alternativa de baixo custo. Os resultados também podem ser usados para desenvolver sistemas automáticos de feedback de treinamento para ajudar os alunos a refletir sobre seu desempenho. Os resultados também podem ter relevância para pesquisadores em áreas como jornalismo ou psicologia social que realizam pesquisas que exigem a avaliação da qualidade da entrevista na mídia, uma vez que isso pode ser feito automaticamente a um custo menor do que o uso de codificadores humanos. Finalmente, os resultados têm o potencial de informar o projeto de sistemas automatizados que podem ser desenvolvidos para ajudar na seleção de pessoal ou na avaliação de funcionários para funções que envolvam a necessidade de se envolver em entrevistas regulares à mídia.

## DECLARAÇÃO DE DISPONIBILIDADE DE DADOS

Os conjuntos de dados apresentados neste estudo podem ser encontrados em repositórios online. Os nomes do repositório e número de acesso podem ser encontrados abaixo: <https://doi.org/10.6084/m9.figshare.11663487>.

## DECLARAÇÃO DE ÉTICA

Os estudos envolvendo participantes humanos foram revisados e aprovados pelo Brunel Research Ethics Office e pelo Comitê de Ética em Pesquisa do Ministério da Defesa. Os participantes forneceram seu consentimento informado por escrito para participar deste estudo.

## CONTRIBUIÇÕES DO AUTOR

MP contribuiu para esta pesquisa conduzindo a pesquisa, incluindo coleta de dados, recrutamento, análise de dados, interpretação e redação inicial deste manuscrito. HM era

## REFERÊNCIAS

- Adams, RB e Kveraga, K. (2015). Visão social: previsão funcional e integração de pistas sociais compostas. *Rev. Filosofia Psychol.* 6, 591–610. doi: 10.1007/s13164-015-0256-1
- Altman, I. (1978). Manwatching: um guia de campo para o comportamento humano. *PsychCRÍTICAS* 23, 533–537.
- Ambady, N., e Rosenthal, R. (1992). Fatias finas de comportamento expressivo como preditores de consequências interpessoais: uma meta-análise. *Psicol. Touro.* 111:256. doi: 10.1037/0033-2909.111.2.256
- Aspegren, K. (1999). Guia BEME No. 2: ensinando e aprendendo habilidades de comunicação em medicina - uma revisão com classificação de qualidade de artigos. *Med. Professor* 21, 563–570. doi: 10.1080/01421599978979
- Babad, E. (1999). Tratamento preferencial em entrevistas na televisão: evidências de comportamento não-verbal. *Comunidade Política.* 16, 337–358. doi: 10.1080/105846099198668
- Bahreini, K., Nadolski, R. e Westera, W. (2016). Rumo ao reconhecimento de emoções de fala em tempo real para e-learning afetivo. *Educ. Informar. Tecnol.* 21, 1367–1386. doi: 10.1007/s10639-015-9388-2
- Bahreini, K., Nadolski, R. e Westera, W. (2017). Treinamento de habilidades de comunicação explorando o reconhecimento multimodal de emoções. *Aprendizado Interativo. Ambiente.* 25, 1065–1082. doi: 10.1080/10494820.2016.1247286
- Barmaki, R. (2016). "Melhorando as habilidades de comunicação social usando o feedback cinético", em *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, (CHI), 86–91.
- Bekele, E., Zheng, Z., Swanson, A., Crittendon, J., Warren, Z. e Sarkar, N. (2013). Compreender como adolescentes com autismo respondem a expressões faciais em ambientes de realidade virtual. *IEEE Trans. Visualização Computação Gráfica* 19, 711–720. doi: 10.1109/TVCG.2013.42
- Bernardini, S., Porayska-Pomsta, K., and Smith, TJ (2014). ECHOES: um jogo sério inteligente para promover a comunicação social em crianças com autismo. *Informar. ciência* 264, 41–60. doi: 10.1016/j.ins.2013.10.027
- Bilakhia, S., Petridis, S., Nijholt, A., and Pantic, M. (2015). O banco de dados de mimetismo MAHNOB: um banco de dados de interações humanas naturalísticas. *Reconhecimento de padrão Deixes* 66, 52–61. doi: 10.1016/j.patrec.2015.03.005
- Boersma, P., e Van Heuven, V. (2001). Fale e desFale com PRAAT. *Glott Inter.* 5, 341–347.
- Chartrand, TL e Bargh, JA (1999). O efeito camaleão: a relação percepção-comportamento e interação social. *J. Pessoa. Sociedade Psicol.* 76, 893. doi: 10.1037//0022-3514.76.6.893
- Chen, C., Lee, I. e Lin, L. (2015). Modelagem autofacial baseada em realidade aumentada para promover a expressão emocional e habilidades sociais de adolescentes com transtornos do espectro do autismo. *Res. Dev. Desativar* 36, 396–403. doi: 10.1016/j.ridd.2014.10.015
- Chen, C., Lee, I. e Lin, L. (2016). Livro de contos de modelagem de vídeo baseado em realidade aumentada de pistas faciais não verbais para crianças com transtorno do espectro do autismo para melhorar suas percepções e julgamentos de expressões faciais e emoções. *Computadores Hum. Behav.* 55, 477–485. doi: 10.1016/j.chb.2015.09.033
- Chen, Z., Wang, L. e Yung, N. (2011). Análise e previsão de movimento humano adaptativo. *Reconhecimento de padrão* 44, 2902–2914. doi: 10.1016/j.patcog.2011.04.022
- Cheng, DS, Salamin, H., Salvagnini, P., Cristani, M., Vinciarelli, A., e Murino, V. (2014). Previsão de classificações de palestras on-line com base em gestos e comportamento vocal. *J. Usuário Multimodal Inter.* 8, 151–160. doi: 10.1007/s12193-013-0142-z
- Chollet, M., Wörtwein, T., Morency, L., Shapiro, A. e Scherer, S. (2015). "Explorando estratégias de feedback para melhorar a fala em público: uma estrutura de audiência virtual interativa", em *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (ACM), 1143–1154. doi: 10.1145/2750858.2806060
- Crane, E., e Gross, M. (2007). "Captura de movimento e emoção: detecção de afeto em movimento de corpo inteiro", em *Proceeding of the International Conference on Affective Computing and Intelligent Interaction*, (Berlin: Springer), 95–101. doi: 10.1007/978-3-540-74889-2\_9
- Curhan, JR e Pentland, A. (2007). Fatias finas de negociação: prever resultados de dinâmicas de conversação nos primeiros 5 minutos. *J. Appl. Psicol.* 92:802.
- Damian, I., Tan, CS, Baur, T., Schöning, J., Luyten, K. e André, E. (2015). "Aumentando as interações sociais: feedback comportamental em tempo real usando técnicas de processamento de sinais sociais", em *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, (ACM), 565–574.
- Darwin, C. (2015). A expressão das emoções no homem e nos animais. Chicago, IL: Imprensa da Universidade de Chicago.
- DeCoster, J. e Claypool, HM (2004). Uma meta-análise dos efeitos de priming na formação de impressões que suportam um modelo geral de vieses informativos. *Pessoa. Sociedade Psicol. Rev.* 8, 2–27. doi: 10.1207/S15327957PSPRO 801\_1
- Deveugele, M. (2015). Treinamento de comunicação: habilidades e além. *Educ. Aconselhamento* 98, 1287–1291. doi: 10.1016/j.pec.2015.08.011
- D'mello, SK e Kory, J. (2015). Uma revisão e meta-análise de sistemas multimodais de detecção de afeto. *Computação ACM. Pesquisas (CSUR)* 47, 1–36. doi: 10.1145/2682899
- Duggan, AP e Parrott, RL (2001). Construção de relacionamento não-verbal dos médicos e fala dos pacientes sobre o componente subjetivo da doença. *Zumbir. Comum. Res.* 27, 299–311. doi: 10.1111/j.1468-2958.2001.tb00783.x
- Dupré, Res. 27, 299–311. doi: 10.1111/j.1468-2958.2001.tb00783.x
- D., Andelic, N., Morrison, G. e McKeown, G. (2018). "Precisão de três sistemas comerciais de reconhecimento automático de emoções em diferentes indivíduos e suas expressões faciais", em *Proceeding of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, (IEEE), 627–632.
- Ekman, P. (2016). Sobre o que os cientistas que estudam as emoções concordam. *Perspectiva. Psicol. ciência* 11, 31–34. doi: 10.1177/1745691615596992
- Ekman, P., e Friesen, WV (1971). Constantes entre culturas no rosto e emoção. *J. Pessoa. Sociedade Psicol.* 17:124. doi: 10.1037/h0030377
- Ekman, P., e Friesen, WV (1986). Uma nova expressão facial pan-cultural de emoção. *Motivação Emo.* 10, 159–168. doi: 10.1007/BF009 92253
- Ekman, PE, Sorenson, R. e Friesen, WV (1969). Elementos pan-culturais em demonstrações faciais de emoção. *Ciência* 164, 86–88. doi: 10.1126/science.164.3875.86

- Ekman, R. (1997). O que a Face Revela: Estudos Básicos e Aplicados da Expressão Espontânea Utilizando o Sistema de Codificação de Ações Faciais (FACS). Oxford: Oxford University Press.
- Esposito, A., Esposito, AM e Vogel, C. (2015). Necessidades e desafios na interação humano-computador para processamento de informações socioemocionais. Carta de Reconhecimento de Padrões. 66, 41–51. doi: 10.1016/j.patrec.2015.02.013 Fiske, J. (2010). Introdução aos Estudos da Comunicação. Milton Park: Routledge.
- Forbes, RJ e Jackson, PR (1980). Comportamento não verbal e o resultado de entrevistas de seleção. J. Psicol Ocupacional. 53, 65–72. doi: 10.1111/j.2044- 8325.1980.tb00007.x
- Frauendorfer, D., Mast, MS, Nguyen, L., e Gatica-Perez, L. (2014). Sensoriamento social não-verbal em ação: registro discreto e extração de comportamento não-verbal em interações sociais ilustrado com um exemplo de pesquisa. J. Comportamento não-verbal. 38, 231–245. doi: 10.1007/s10919-014-0173-5
- Gao, Y., Bianchi-Berthouze, N. e Meng, H. (2012). O que o toque nos diz sobre emoções na jogabilidade baseada em tela sensível ao toque? ACM Trans. Computador-Hum. Interação (TOCHI) 19, 1–30. doi: 10.1145/2395131.2395138
- Giles, H., Klaus RS e Taylor, DM (1979). "Marcadores de fala na interação social," em Social Markers in Speech, eds. KR Scherer e H. Giles (Cambridge, Reino Unido: Cambridge University Press), 343–381.
- Glass, CR, Merluzzi, TV, Biever, JL e Larsen, KH (1982). Avaliação cognitiva da ansiedade social: desenvolvimento e validação de um questionário de autoafirmação. cogn. Terapia Res. 6, 37–55. doi: 10.1007/BF01185725 Goldin-Meadow, S., e Alibali, MW (2013). O papel do gesto na fala, aprendizagem, e criando linguagem. Ana. Rev. Psychol. 64, 257-283.
- Goutte, C., e Gaussier, E. (2005). "Uma interpretação probabilística de precisão, recordação e F-score, com implicações para avaliação," em Proceeding of the European Conference on Information Retrieval, (Berlin: Springer), 345–359. doi: 10.1007/978-3-540-31865-1\_25
- Gross, JJ e Feldman-Barrett, L. (2011). Geração de emoções e regulação de emoções: um ou dois depende do seu ponto de vista. Emoção Rev. 3, 8–16.
- doi: 10.1177/1754073910380974
- Gross, MM, Crane, EA e Fredrickson, BL (2012). Esforço-forma e avaliação cinemática da expressão corporal da emoção durante a marcha. Zumbir. Jogada. ciência 31, 202–221. doi: 10.1016/j.humov.2011.05.001
- Hall, JA, Horgan, TG e Murphy, NA (2019). Comunicação não verbal. Ana. Rev. Psychol. 70, 271–294.
- Han, Y., e Nunes, J. (2010). "Leia o sinal, mas não o mencione: como o consumo conspicuo embaraça o sinalizador", em NA - Advances in Consumer Research, vol. 37, eds MC Campbell, J. Inman e R. Pieters (Duluth, MN: Association for Consumer Research), 81–84.
- Hart, Y., Czerniak, E., Karnieli-Miller, O., Mayo, AE, Ziv, A., Biegon, A., et al. (2016). Análise automatizada de vídeo de comunicação não-verbal em um ambiente médico. Frente. Psicol. 7:1130. doi: 10.3389/fpsyg.2016.01130
- Hashimoto, M., Yamano, M. e Usui, T. (2009). "Efeitos da sincronização emocional nas comunicações humano-robô KANSEI," em Proceeding of the RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication, (IEEE), 52–57.
- Hill, CE, Siegelman, L., Gronsky, BR, Sturniolo, F. e Fretz, BR (1981). Comunicação não verbal e resultado do aconselhamento. J. Aconselhamento Psychol. 28:203. doi: 10.1037/0022-0167.28.3.203
- Ho, S., Foulsham, T. e Kingstone, A. (2015). Falar e ouvir com os olhos: sinalização do olhar durante interações diádicas. PLoS One 10:e0136905. doi: 10.1371/journal.pone.0136905
- Holding, BC, Sundelin, T., Lekander, M. e Axelsson, J. (2019). Privação do sono e seus efeitos na comunicação durante tarefas individuais e colaborativas. ciência Rep. 9, 1–8. doi: 10.1038/s41598-019-3 9271-6
- Hunyadi, L. (2019). Concordar/discordar em um diálogo: padrões multimodais de sua expressão. Frente. Psicol. 10:1373. doi: 10.3389/fpsyg.2019.01373 Kim, M. e Suzuki, K. (2014). Estudo comparativo do comportamento humano no jogo de cartas com um companheiro humanoíde. Int. J. Soc. Robô. 6, 5–15. doi: 10.1007/ s12369-013-0184-0
- Knapp, ML, Hall, MA e Horgan, TG (2013). Comunicação não verbal em Interação humana. Boston: Cengage Learning.
- Knutson, B. (1996). Expressões faciais de emoção influenciam inferências de traços interpessoais. J. Comportamento não-verbal. 20, 165–182. doi: 10. 1007/BF02281954
- Koskimäki, H., Mönttinen, H., Siirtola, P., Huttunen, HL, Halonen, R. e Rönning, J. (2017). "Detecção precoce de ataques de enxaqueca com base em sensores vestíveis: experiências de coleta de dados usando o Empatica E4", em Anais da Conferência Conjunta Internacional ACM 2017 sobre Computação Pervasiva e Ubíqua e Anais do Simpósio Internacional ACM 2017 sobre Computadores vestíveis (ACM), 506–511.
- Lacerda, F. (2009). LVA-tecnologia—A ilusão de "detecção de mentiras". Em FONETIK. Disponível online em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.160.7742&rep=rep1&type=pdf#page=220> (acessado em 7 de janeiro de 2018).
- Lapidot-Leffer, N., e Barak, A. (2012). Efeitos do anonimato, invisibilidade e falta de contato visual na desinibição tóxica online. Computadores Hum. Behav. 28, 434–443. doi: 10.1016/j.chb.2011.10.014
- Lieberman, P. (1976). Sistemas prosódicos e entonação em inglês. Idioma 52, 508–511.
- Liu, C., Calvo, RA e Lim, R. (2016). Melhorando a consciência dos estudantes de medicina sobre sua comunicação não-verbal por meio de comportamento não-verbal automatizado retorno. Frente. TIC 3:11. doi: 10.3389/fict.2016.00011
- Macmillan, NA e Creelman, DC (2004). Teoria da Detecção: Um Guia do Usuário. Londres: Imprensa de psicologia.
- Mandrek, JN (2011). Medidas de acordo interintermediário. J. Thoracic Oncol. 6, 6–7. doi: 10.1097/JTO.0b013e318200f983
- Mayew, VJ e Venkatachalam, M. (2010). Apêndice da Internet para "O poder da voz: estados afetivos gerenciais e desempenho futuro da empresa. Disponível online em: <https://faculty.fuqua.duke.edu/~{}vmohan/bio/files/published%20papers/InternetAppendixMV2011.pdf> (acessado em 12 de dezembro de 2020).
- Meissel, K., Meyer, F., Yao, ES e Rubie-Davies, CM (2017). Subjetividade dos julgamentos do professor: explorando as características dos alunos que influenciam os julgamentos do professor sobre a capacidade do aluno. Ensinar. Professor Educ. 65, 48–60.
- Naim, I., Tanveer, I., Gildea, D. e Hoque, ME (2016). Análise automatizada e previsão de desempenho em entrevistas de emprego. IEEE Trans. Computação Afetiva. 9, 191–204. doi: 10.1109/TAFFC.2016.2614299
- Nguyen, LS, e Gatica-Perez, D. (2015). "Eu contrariaria você em um minuto: fatias finas de comportamento não verbal em entrevistas de emprego", em Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, (ACM), 51–58.
- Olguin, DO e Pentland, AS (2007). "Emblemas sociométricos: estado da arte e aplicações futuras", em Anais do Colóquio de Doutorado apresentado no IEEE 11º Simpósio Internacional de Computadores vestíveis, (Boston, MA: IEEE).
- Ortony, A., e Turner, TJ (1990). O que há de básico nas emoções básicas? Psicol. Apocalipse 97:315. doi: 10.1037/0033-295X.97.3.315
- Pantic, M., Cowie, R., D'Errico, F., Heylen, D., Mehu, M., Pelachaud, C., et al. (2011). "Processamento de sinal social: a agenda de pesquisa," em Visual Analysis of Humans, eds T. Moeslund, A. Hilton, V. Krüger e L. Sigal (Londres: Springer). doi: 10.1007/978-0-85729-997-0\_26
- Pantic, M., Sebe, N., Cohn, JC e Huang, T. (2005). "Interação humano-computador multimodal afetiva," em Proceedings of the 13th Annual ACM International Conference on Multimedia, (ACM), 669–676.
- Park, S., Shim, HS, Chatterjee, M., Sagae, K. e Morency, P. (2016). Análise multimodal e previsão de persuasão em mídias sociais online. ACM Trans. Interagir. Sistema Inteligente (TIIS) 6, 1–25. doi: 10. 1145/2897739
- Paxton, A., Rodriguez, K. e Dale, R. (2015). PsyGlass: aproveitando o Google Glass para coleta de dados naturalísticos. Behav. Res. Métodos 47, 608–619. doi: 10.3758/s13428-015-0586-z
- Pentland, A. e Heibeck, T. (2010). Sinais honestos: como eles moldam nosso mundo. Cambridge: Imprensa do MIT.
- Pereira, M., e Hone, K. (2021). "Intervenção de treinamento de habilidades de comunicação com base no reconhecimento automatizado de sinais não-verbais", em Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, (CHI), 1–14.
- Pollick, FE, Paterson, HE, Bruderlin, A. e Sanford, AJ (2001). Percebendo o afeto do movimento do braço. Cognição 82, B51–B61. doi: 10.1016/S0010-0277 (01)00147-0 Potamianos, A. (2014). "Processamento multimodal cognitivo: do sinal ao comportamento", em Anais do Workshop de 2014 sobre o mapeamento do futuro da pesquisa de interação multimodal, incluindo oportunidades e desafios de negócios, (ACM), 27–34. doi: 10.1145/2666253.2666264
- Ramdinmawii, E., Mohanta, A., e Mittal, VK (2017). "Reconhecimento de emoção do sinal de fala", em Proceedings of the TENCON 2017-2017 IEEE Region 10 Conference, (Piscataway, NJ: IEEE), 1562–1567.

Rasipuram, S., e Jayagopi, DB (2018). avaliação automática do comunicação habilidade de baseado em entrevista interações. Aplicativo de ferramentas multimídia. 77, 18709–18739. doi: 10.1007/s11042-018-5654-9

Rozin, P. e Cohen, AB (2003). Alta frequência de expressões faciais correspondentes a confusão, concentração e preocupação em uma análise de expressões faciais naturais de americanos. Emoção 3:68. doi: 10.1037/1528-3542.3.1.68 Schefflen, AE (1964). O significado da postura nos sistemas de comunicação.

Psiquiatria 27, 316-331. doi: 10.1080/00332747.1964.11023403  
Sondhi, S., Khan, M., Vijay, R. e Salhan, AK (2015). Indicadores vocais de estresse emocional. Int. J. Comput. Appl. 122, 38-43.  
Spitzberg, BH e Adams, TW (2007). CSRS, a Escala de Avaliação de Habilidades de Conversação: uma Avaliação Instrucional de Competência Interpessoal. Washington, DC: Associação Nacional de Comunicação.  
Stöckli, S., Schulte-Mecklenbeck, M., Borer, S. e Samson, AC (2018). Análise de expressão facial com AFFDEX e FACET: um estudo de validação. Behav. Res. Métodos 50, 1446–1460. doi: 10.3758/s13428-017-0996-1  
Sung, M. e Pentland, A. (2005). PokerMetrics: Detecção de Estresse e Mentira Através de Sensoriamento Fisiológico Não Invasivo. Relatório técnico. Cambridge, MA: MIT Media Lab.

Sullivan, J. (2018). O efeito de primazia na formação de impressões: algumas replicações e extensões. Sociedade Psicol. Pessoa. ciência 10, 432-439. doi: 10.1177/1948550618771003  
Tao, J. e Tan, T. (2009). Processamento Afetivo da Informação. Londres: Springer.  
Taylor, R. (2015). Técnicas de entrevista na mídia: um guia completo para a mídia Treinamento. Londres: Kogan Page Publishers.  
Turk, M. (2014). Interação multimodal: uma revisão. pat. Reconhecer. Deixe 36, 189–195. doi: 10.1016/j.patrec.2013.07.003 Usui, T., Kume, K., Yamano, M. e Hashimoto, M. (2008). "Um sistema robótico de comunicação KANSEI baseado em sincronização emocional," em Proceeding of the 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems, (IEEE), 3344–3349.

Van Lierde, K., Van Heule, S., De Ley, S., Mertens, E. e Claeys, S. (2009). Efeito do estresse psicológico na qualidade vocal feminina. Folia Phoniatr. Logopaed. 61, 105–111.

Vinciarelli, A., Pantic, M. e Bourlard, H. (2009). Processamento de sinais sociais: levantamento de um domínio emergente. Computação de visão de imagem. 27, 1743–1759. doi: 10.1016/j.imavis.2008.11.007  
Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Preenchendo a lacuna entre o animal social e a máquina não social: uma pesquisa sobre o processamento de sinais sociais. Transação IEEE. Afetar. Comput. 3, 69-87.  
Vrij, A., Edward, K., Roberts, KP e Bull, R. (2000). Detectando o engano por meio da análise do comportamento verbal e não verbal. J. Comportamento não-verbal. 24, 239–263. doi: 10.1023/A:1006610329284  
Witten, IH e Frank, E. (2002). Mineração de dados: ferramentas e técnicas práticas de aprendizado de máquina com implementações Java. ACM Sigmod. Registro. 31, 76–77. doi: 10.1145/507338.507355  
Zhang, Y., Olenick, J., Chang, C., Kozlowski, S. e Hung, H. (2018a). TeamSense: avaliando o afeto pessoal e a coesão do grupo em pequenas equipes por meio de interação diádica e análise de comportamento com sensores vestíveis. Proc. ACM Inter. Mob. Tecnologia Ubiqua Vestível. 2, 1–22. doi: 10.1145/3264960 Zhang, Y., Olenick, J., Chang, CH, Kozlowski, S. e Hung, H. (2018b). "The i in team: Mineração de rotina de interação social pessoal com modelos de tópicos de dados de equipe de longo prazo", em Proceeding of the 23rd International Conference on Intelligent User Interfaces, (ACM), 421–426.

**Conflito de Interesses:** Os autores declaram que a pesquisa foi conduzida na ausência de quaisquer relações comerciais ou financeiras que possam ser interpretadas como um potencial conflito de interesses.

**Nota do editor:** todas as reivindicações expressas neste artigo são exclusivamente dos autores e não representam necessariamente as de suas organizações afiliadas ou do editor, dos editores e dos revisores. Qualquer produto que possa ser avaliado neste artigo, ou reclamação que possa ser feita por seu fabricante, não é garantido ou endossado pelo editor.

Copyright © 2021 Pereira, Meng e Hone. Este é um artigo de acesso aberto distribuído sob os termos da Creative Commons Attribution License (CC BY). O uso, distribuição ou reprodução em outros fóruns é permitido, desde que o(s) autor(es) original(is) e o(s) detentor(es) dos direitos autorais sejam creditados e que a publicação original nesta revista seja citada, de acordo com a prática acadêmica aceita. Nenhum uso, distribuição ou reprodução é permitido que não esteja de acordo com estes termos.