

Veja discussões, estatísticas e perfis de autores para esta publicação em: <https://www.researchgate.net/publication/331160481>

WIYE: construindo um corpus de gravações de áudio e vídeo de crianças com um aplicativo baseado em histórias

Documento da Conferência · Fevereiro de 2019  
DOI: 10.1145/3308557.3308684

CITAÇÕES  
2

LÊ  
119

4 autores:



Micol Spitale

Universidade de Cambridge

31 PUBLICAÇÕES 118 CITAÇÕES

VER PERFIL



Fabio Catânia

Politécnico de Milão

21 PUBLICAÇÕES 108 CITAÇÕES

VER PERFIL



Giulia Cosentino

Politécnico de Milão

20 PUBLICAÇÕES 72 CITAÇÕES

VER PERFIL




Mirko Gelsomini


Politécnico de Milão

64 PUBLICAÇÕES 720 CITAÇÕES

VER PERFIL

Alguns dos autores desta publicação também estão trabalhando nesses projetos relacionados:

- 

Projeto M4All - Motion 4 All [View](#)
- 

MOODY @ MIT [Ver projeto](#)

## WIYE: construindo um corpus de gravações de áudio e vídeo de crianças com um aplicativo baseado em histórias

Micol Spitale  
micol.spitale@polimi.it  
Politecnico di Milano  
Milão, Itália

Fabio Catania  
fabio.catania@polimi.it  
Politecnico di Milano  
Milão, Itália

Giulia Cosentino  
giulia.cosentino@polimi.it  
Politecnico di Milano  
Milão, Itália

Mirko Gelsomini  
mirko.gelsomini@polimi.it  
Politecnico di Milano  
Milão, Itália

Franca Garzotto  
franca.garzotto@polimi.it  
Politecnico di Milano  
Milão, Itália

### ABSTRATO

Este artigo descreve o procedimento para a criação de um conjunto de dados emocionais, denominado WIYE (What Is Your Emotion?), composto por conteúdos semânticos, gravações de áudio e vídeo de crianças. Os dados foram coletados por meio de um aplicativo interativo de contação de histórias, levando crianças de 4 a 12 anos a descobrir sobre sua esfera emocional e habilidades de expressão de emoções. Durante a história, cada episódio é dedicado e focado em uma emoção específica. Os estados emocionais investigados são tristeza, raiva, medo, surpresa e alegria. Este corpus pode ser explorado por Tecnologias Conversacionais: pode ser usado com algoritmos de classificação de Aprendizado de Máquina para treinar modelos para reconhecer emoções expressas por crianças a partir do tom de sua voz, sua expressão facial e o conteúdo de suas conversas.

### CONCEITOS DE CCS

- **Computação centrada no ser humano** e Interfaces de linguagem natural.

### PALAVRAS-CHAVE

Discurso emocional e corpus facial; Banco de dados emocional; Computação Afetiva; Reconhecimento de Emoções; Classificação da emoção.

#### Formato de Referência ACM:

Micol Spitale, Fabio Catania, Giulia Cosentino, Mirko Gelsomini e Franca Garzotto. 2019. WIYE: construindo um corpus de gravações de áudio e vídeo infantil com um aplicativo baseado em histórias. Na 24ª Conferência Internacional sobre Interfaces de Usuário Inteligentes (IUI '19 Companion), 17 a 20 de março de 2019, Marina del Rey, CA, EUA. ACM, Nova York, NY, EUA, 2 páginas. <https://doi.org/10.1145/3308557.3308684>

### 1. INTRODUÇÃO

A comunicação consiste na transmissão de informações verbais e não verbais. Segundo [2], as expressões faciais são responsáveis

para cerca de 55% da percepção da mensagem, entonação de voz para cerca de 38% e palavras reais apenas para 7%. Isso enfatiza a importância da capacidade de expressar e entender informações não comunicadas abertamente como um conteúdo. As emoções geralmente estão entre esse tipo de informação. **Uma emoção é um episódio de mudanças coordenadas em vários componentes, incluindo pelo menos ativação neurofisiológica, expressão motora e sentimento subjetivo, mas possivelmente também tendências de ação e processos cognitivos em resposta a eventos externos ou internos de grande importância para o organismo**. Na psicologia, existem muitas teorias que fornecem diferentes maneiras de organizar as emoções nas aulas. Uma das mais populares é a de Ekman [3], que identificou seis emoções básicas que podem ser universalmente reconhecidas: alegria, tristeza, surpresa, medo, raiva e nojo. Estes são chamados os Seis Grandes. O impacto das teorias da emoção em outras áreas de pesquisa, como computação afetiva na ciência da computação, é extraordinariamente alto. Por esta razão, nos últimos anos tem-se pesquisado muito para possibilitar a interação humano-máquina verbal e visual emocional, desenvolver sistemas de segurança, aumentar a segurança do motorista e investigar no campo da saúde. Para atingir todos esses objetivos ambiciosos, um banco de dados emocional contendo gravações de áudio e vídeo rotuladas é um pré-requisito. Apesar de sua utilidade, os conjuntos de dados emocionais atualmente disponíveis são limitados, pois a maioria deles captura apenas as expressões emocionais (com o rosto e a voz) em adultos.

Neste artigo, descrevemos o processo de desenvolvimento do WIYE, nosso banco de dados emocional original que armazena arquivos de áudio e vídeo, imagens e enunciados de crianças italianas de 4 a 12 anos. WIYE é o primeiro corpus italiano que coleta arquivos multimídia explorando canais de comunicação visuais, orais e semânticos. Além disso, é o primeiro conjunto de dados italiano de fala emocional de crianças e a primeira coleção de enunciados com conteúdo emocional produzidos com a contribuição direta de jovens. Este corpus tem uma ampla gama de aplicações possíveis: por exemplo, pode ser usado para treinar algoritmos de Machine Learning (ML) em Speech-to-Text a partir de voz emocional infantil e em reconhecimento e síntese automática de emoções a partir da voz e a expressão facial. Este tipo de sistema tem sido muito investigado recentemente como ferramenta didática de apoio e no estudo de Distúrbios do Neurodesenvolvimento (NDD) como o Transtorno do Espectro Autista [8], [4].

A permissão para fazer cópias digitais ou impressas de todo ou parte deste trabalho para uso pessoal ou em sala de aula é concedida sem taxa, desde que as cópias não sejam feitas ou distribuídas com fins lucrativos ou vantagens comerciais e que as cópias contenham este aviso e a citação completa na primeira página. Os direitos autorais de componentes deste trabalho pertencentes a outros que não a ACM devem ser respeitados. Abstraindo com crédito é permitido. Para copiar de outra forma, ou republicar, postar em servidores ou redistribuir para listas, requer permissão específica prévia e/ou taxa. Solicite permissões de [permissions@acm.org](mailto:permissions@acm.org).  
IUI '19 Companion, 17 a 20 de março de 2019, Marina del Rey, CA, EUA © 2019 Association for Computing Machinery.  
ACM ISBN 978-1-4503-6673-1/19/03... \$ 15,00  
<https://doi.org/10.1145/3308557.3308684>

## 2 WIYE: O CORPUS

O corpus WIYE é um banco de dados emocional que armazena gravações de áudio e vídeo, fotos e enunciados de crianças italianas de 4 a 12 anos. A versão atual do conjunto de dados conta com 710 gravações de áudio e vídeo e 710 enunciados originais com semântica emocional de 142 crianças divididas igualmente entre homens e mulheres.

As gravações e enunciados foram coletados seguindo um procedimento sistemático dentro de um jogo, solicitando às crianças que agissem expressando emoções com a voz e com o rosto e, em seguida, inventassem frases com conteúdos emocionais. **As emoções que investigamos são as Big Six excluindo o nojo**, como faz o indico.io [1], que realiza o reconhecimento de emoções a partir de texto e expressão facial. Toda a experiência do usuário foi pensada como um jogo para que as crianças se sintam confortáveis durante as gravações. Em particular, o jogo é um aplicativo baseado em histórias que explora a Tecnologia Conversacional: é um Agente Conversacional (CA), ou sistema de diálogo, ou seja, um programa de software capaz de interagir com seres humanos por meio de linguagem natural. O aplicativo foi projetado de acordo com o padrão educacional graças à cooperação com psicólogos e especialistas em desenvolvimento infantil que criaram o conteúdo. De fato, a narração é inspirada em um romance clássico contado na escola primária [5] adaptado a um ambiente espacial e estruturado em diferentes cenas.

### 2.1 O cenário Antes

de iniciar o jogo, os pais dos participantes nos deram autorização para gravar seus filhos e recebemos o IRB

aprovação da universidade. Os pequenos participantes foram convidados a entrar um a um em uma tenda espacial e mergulhar em um planeta diferente. O cenário era composto por algumas almofadas onde eles podiam sentar e um tablet, fones de ouvido e um microfone para isolá-los do ruído.

### 2.2 A sessão de contação de histórias O

**personagem principal da história é um alienígena chamado Boo que pede ao usuário que o ensine a expressar emoções corretamente tanto com a voz quanto com as expressões faciais.** De fato, Boo pode se comunicar apenas com entonação neutra e sem manifestação de emoção facial.

**A história é dividida em cinco cenas diferentes**, cada uma correspondendo a uma emoção. Durante cada cena, o usuário é convidado a realizar três tarefas diferentes:

- 1) **repetir uma frase específica**, já soletrada pelo alienígena durante a contação da história. O enunciado precisa ser expresso com a emoção adequada e especificada. As frases atribuídas foram escolhidas com uma semântica emocional para que as crianças entrem no mesmo humor do personagem principal; 2) **inventar uma frase ligada ao clima atual da história.**

Para isso, pede-se que usem a imaginação e a inteligência emocional, bem como a capacidade de compreensão do contexto;

- 3) **mostrar sua expressão facial condizente com aquela emoção.**

Para todas essas tarefas, a conselho dos especialistas, confiamos na capacidade de auto-indução das crianças, imaginando ou lembrando uma situação em que a emoção desejada foi sentida fortemente, o que é conhecido como método de Stanislavski [6].

### 2.3 O método de armazenamento

Coletamos uma gravação em vídeo e uma foto para cada emoção expressa com o rosto e, em seguida, dez arquivos de áudio com a voz de cada participante: cinco para as frases a serem repetidas e cinco para as inventadas. No banco de dados, nenhum dado armazenado é atribuído diretamente a um usuário específico.

## 3 CONCLUSÃO E TRABALHOS FUTUROS

Este artigo apresentou uma descrição detalhada do corpus emocional WIYE, que armazena gravações de áudio e vídeo, fotos e enunciados de crianças italianas de 4 a 12 anos de idade. Atualmente, o conjunto de dados contabiliza 710 gravações de áudio e 710 vídeos e 710 enunciados com semântica emocional de 142 crianças divididas igualmente entre homens e mulheres.

Embora as expressões encenadas tenham limitações intrínsecas relacionadas à sua naturalidade em relação às espontâneas, acreditamos que o WIYE é adequado para uso em Speech-to-Text a partir de voz emocional infantil e no reconhecimento e síntese automática de emoções. Com efeito, a base de dados está prevista para ser utilizada para o desenvolvimento de novos algoritmos para o reconhecimento de expressões emocionais vocais e faciais por crianças e para análise emocional cross-modal, comparando o tom de áudio, a semântica da fala, a expressão facial e dados biológicos. Uma limitação do método de contação de histórias usado para coletar sentenças emocionais produzidas pelos usuários é que as crianças foram solicitadas a inventar sentenças vinculadas a contextos pré-definidos e isso produziu um conjunto de enunciados topicamente homogêneo. Isso tem implicações enormes e negativas no treinamento do modelo, uma vez que o modelo seria facilmente superajustado se o corpus não fosse diversificado o suficiente. O seguimento natural deste trabalho é a ampliação do primeiro conjunto de dados preliminares dentro de uma experiência já planejada para o ensino fundamental. Na verdade, estamos trabalhando para refinar o aplicativo baseado em história para garantir uma experiência de jogo mais envolvente e envolvente. Isso pode ser explorado para coletar uma quantidade maior de dados, útil para muitas aplicações, incluindo pesquisas voltadas para crianças para investigar melhor as habilidades e métodos de comunicação verbal e não verbal das crianças. Paralelamente, também estamos trabalhando no processo de validação via crowdsourcing para aumentar a qualidade das gravações e enunciados fornecidos.

## AGRADECIMENTOS

Somos muito gratos pela cooperação aos psicólogos, aos cuidadores, às crianças e seus pais.

## REFERÊNCIAS

- [1] 2016. Indico - Reconhecimento de emoções. <https://indico.io> (2016).
- [2] Claude C Chibielushi et al. 2003. Reconhecimento de expressão facial: Uma breve visão geral do tutorial. CVonline: Compêndio On-Line de Visão Computacional (2003).
- [3] Paul Ekman. 1992. Um argumento para emoções básicas. Cognição e emoção (1992).
- [4] Liliana Laranjo et al. 2018. Agentes conversacionais em saúde: uma revisão sistemática. Jornal da Associação Americana de Informática Médica (2018).
- [5] David McKee et al. 1990. Elmer: o elefante variopinto. Mondadori.
- [6] Perviz Sawoski. 2010. O Sistema Stanislavski: crescimento e metodologia. Colégio Santa Mônica (2010).
- [7] Klaus R Scherer. 2003. Comunicação vocal de emoção: uma revisão de pesquisa paradigmática. Comunicação de fala (2003).
- [8] Hiroki Tanaka et al. 2017. Agentes conversacionais incorporados para treinamento multimodal automatizado de habilidades sociais em pessoas com transtornos do espectro do autismo. PloS um (2017).