

Eye movements and human face perception: An holistic analysis and proficiency classification based on frontal 2D face images

Victor P. L. Varela¹, Estela Ribeiro¹, Pedro A. S. S. Orona¹, Carlos E. Thomaz¹

¹Departamento de Engenharia Elétrica – Centro Universitário FEI (FEI)
Avenida Humberto Alencar Castelo Branco – São Bernardo do Campo – SP – Brazil

{vvarela, cet}@fei.edu.br

Abstract. *Human faces convey a collection of information, such as gender, identity, and emotional states. Therefore, understanding the differences between volunteers' eye movements on benchmark tests of face recognition and perception can explicitly indicate the most discriminating regions to improve performance in this visual cognitive task. The aim of this work is to qualify and classify these eye strategies using multivariate statistics and machine learning techniques, achieving up to 94.8% accuracy. Our experimental results show that volunteers have focused their visual attention, on average, at the eyes, but those with superior performance in the tests carried out have looked at the nose region more closely.*

Resumo. *Faces humanas transmitem uma coleção de informações, tais como gênero, identidade e estados emocionais. Portanto, compreender as diferenças entre os movimentos oculares de voluntários em testes consolidados da literatura sobre reconhecimento e percepção faciais podem indicar explicitamente as regiões mais discriminantes para melhor desempenho nessa tarefa cognitiva visual. O objetivo desse artigo é qualificar e classificar essas estratégias oculares utilizando técnicas de estatística multivariada e de aprendizado de máquina, atingindo acurácia de até 94,8%. De acordo com a base analisada, voluntários observaram, em média, os olhos, porém, aqueles com rendimento superior nos testes, tenderam a observar mais atentamente a região do nariz.*

1. Introdução

A maioria das pessoas possui aproximadamente o mesmo nível de experiência quando o assunto é reconhecimento facial, entretanto existem consideráveis diferenças nas nossas habilidades individuais em reconhecer faces [Bate et al. 2010]. Essas diferenças podem ser descritas entre os que possuem habilidades superiores, chamados Super-Reconhecedores (SR) [Russell et al. 2009, Robertson et al. 2016, White et al. 2015], até os que são afetados por prosopagnosia [Bodamer 1947]. As pessoas desse último grupo sofrem de severos problemas em reconhecer faces, que podem ocorrer devido a danos ou doenças neurológicas, deficiências visuais, intelectuais e até mesmo dificuldades sócio-emocionais [Bate and Cook 2012, Jones and Tranel 2001].

Características faciais são, sem dúvida, o que nos fazem ser únicos no mundo e, por conta disso, faz-se importante entender porque a maioria das pessoas possui a capacidade de identificar instantaneamente seus familiares e amigos, por exemplo, mas demonstram menos facilidade para rostos desconhecidos.

Pesquisas mostram que o processamento facial não necessariamente é estruturado de maneira simples e lógica, porém sugerem que existem padrões nas estratégias oculares para codificação, integração e interpretação de faces [Bruce and Young 2012, Grüter et al. 2008, Gobbini and Haxby 2007, Thomaz et al. 2017].

O objetivo desse artigo é investigar e revelar, dentre voluntários comuns, as diferenças nos padrões de estratégia ocular em um teste internacional de percepção facial de referência da literatura afim [Burton et al. 2010]. Dividindo esses voluntários conforme suas habilidades, foi possível observar essas diferenças e, de forma inédita, classificá-las linearmente com acurácia de até 94,8%. Os resultados mostram que os voluntários, em média, e de acordo com a base de imagens de faces analisada, observaram os olhos do indivíduo, porém, aqueles com rendimento superior, tenderam a observar mais atentamente a região do nariz.

Este presente artigo está dividido em 4 seções, como se segue. Na seção 2, descrevem-se os materiais e métodos utilizados para execução do experimento. Na seção 3, relatam-se os resultados obtidos nas análises visuais e de classificação utilizando as técnicas de estatística multivariada PCA e MLDA, descritas na seção 2. Por fim, na seção 4 tem-se a conclusão deste trabalho e possibilidades de trabalhos futuros.

2. Materiais e Métodos

2.1. Participantes

O experimento contou com a participação de 39 voluntários (25 mulheres e 16 homens) entre 13 e 46 anos ($M = 17,63$ anos; $DP = 6,25$). Foram utilizados apenas participantes com acuidade de visão normal. Todos assinaram um termo de consentimento livre e esclarecido (TCLE) e participaram de uma entrevista em que responderam as seguintes questões:

- Idade
- Sexo
- Formação Acadêmica
- Usuário de óculos? (SIM ou NÃO)
- Tempo estimado desde a última refeição
- Sente cansaço físico? (SIM ou NÃO)
- Sente cansaço mental? (SIM ou NÃO)
- Fez uso de medicamento no dia do experimento? (SIM ou NÃO) Qual?
- Fumante? (SIM ou NÃO)
- Diagnosticado previamente com prosopagnosia? (SIM ou NÃO)

2.2. Aquisição de sinal e equipamento

A aquisição do sinal foi feita em uma sala fechada, com iluminação artificialmente controlada dentro das especificações ideais entre 300 e 1000 lux e posicionada fora do campo visual do participante. Foi utilizado um equipamento de rastreamento ocular da marca Tobii modelo TX300 com capacidade de captura de dados de 300 Hz, um teclado comum padrão ABNT e um notebook com processador core i7 e 16Gb de RAM, sistema operacional Windows 7 e um software de apoio para a calibração do equipamento e mineração dos dados coletados. O equipamento de rastreamento ocular utilizado é composto por um

monitor TFT de 23 polegadas com sistema de iluminação infravermelho e 2 câmeras de captura acopladas na parte inferior. Antes de iniciar o experimento, todos participantes foram acomodados em frente ao equipamento de rastreamento ocular e orientados a olhar apenas para a tela a sua frente, evitando fechar os olhos ou fazer movimentos bruscos com a cabeça para não perder o sinal. Com o auxílio do software aplicativo Tobii Studio, o equipamento foi calibrado a cada coleta, ajustando a posição do participante dentro do alcance do equipamento e o foco da fixação ocular através de uma grade de calibração.

2.3. Teste de percepção facial

Utilizou-se o Glasgow Face Matching Test (GFMT), versão curta [Burton et al. 2010]. Esse teste tem sido usado recentemente para estudo de super-reconhecedores [Robertson et al. 2016]. É um teste que usa de aspectos de percepção na correlação de imagens de faces. A ideia inovadora proposta por [Burton et al. 2010] foi a de utilizar duas fotos do mesmo indivíduo, porém uma tirada após algum tempo depois da anterior, e com outra câmera. Esse teste é considerado difícil e foi criado para complementar outros testes existentes de processamento facial (e.g. CFMT [Duchaine and Nakayama 2006]). Por utilizar câmeras diferentes, tenta imitar situações do dia-a-dia de equipes de segurança em fazer a correlação de rostos não familiares em imagens semelhantes.

2.4. Procedimento experimental

Fora explicado previamente aos participantes o objetivo do presente estudo e feita uma breve descrição acerca do equipamento de rastreamento ocular. Caso aceitassem em participar da pesquisa, assinaram o TCLE, responderam à entrevista e foram posteriormente posicionados frente ao equipamento de rastreamento ocular. Cada participante realizou o experimento individualmente sob as instruções e companhia do pesquisador, em conjunto das descrições do experimento no monitor do equipamento.

A tarefa consistiu em, enquanto o equipamento capturava os movimentos oculares, o participante expor em voz alta a sua resposta perante cada par de imagem do teste GFMT [Burton et al. 2010] (de um total de 40) e, como definido em Burton et al. (2010), responder se o par de rosto apresentado tratava-se do mesmo indivíduo ou não.

2.5. Análise de proficiência

O GFMT [Burton et al. 2010] foi inicialmente proposto para ser um teste em que a proficiência do voluntário é medida apenas por quantidade de acertos, porém, observado em [Sergent and PONCET 1990], incluir o tempo utilizado para realização do teste, em adição à precisão, é tido como uma boa métrica para análise de controles. Portanto, para discriminar ainda mais os grupos de acordo com sua proficiência, foi utilizado o método de cálculo para pontuação individual proposto por Volke [Volke et al. 2002], que inclui essas duas variáveis (assertividade e tempo), como ilustra a Equação 1:

$$Hs = (N_{correct} - \frac{N}{2}) \times \frac{RTm}{RTs}, \quad (1)$$

onde Hs é a proficiência de cada voluntário (a ser calculada); $N_{correct}$ é o número de questões respondidas corretamente pelo voluntário, providas do teste GFMT [Burton et al. 2010]; N o número total de questões (40); RTm é o tempo de resposta médio de todos os participantes em todas as questões; RTs é o tempo médio de resposta

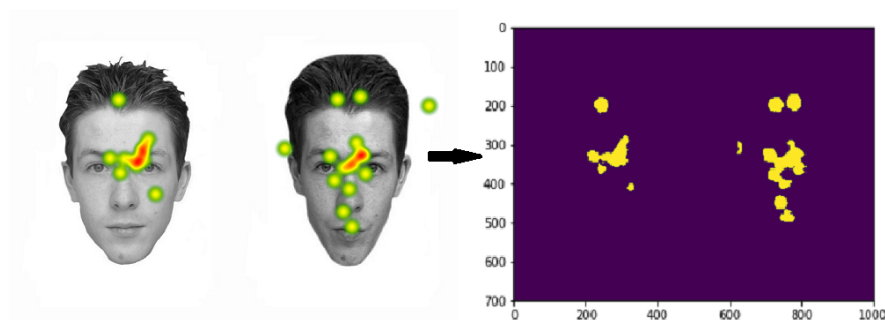


Figure 1. Exemplo de binarização de mapa de olhar em uma imagem.

de cada voluntário da pesquisa. É importante notar que voluntários que respondam todas as perguntas com a mesma alternativa procurando alcançar no pior caso 50% de acerto, mas no tempo mínimo possível, ficam com pontuação igual a zero [Volke et al. 2002].

Os participantes foram separados em ordem de acertos, e o coeficiente de Volke [Volke et al. 2002] foi utilizado como critério de desempate entre grupos de mesma quantidade de acertos. Após esse procedimento os participantes foram divididos em 4 quartis de proficiência, sendo o quartil 1 responsável pelos 10 piores (por quantidade de acerto) e o quartil 4 os 10 melhores (por quantidade de acerto), o quartil 2 contou com 9 participantes e o quartil 3, 10.

2.6. Classificação dos movimentos oculares

Com os resultados obtidos pelo equipamento de rastreamento ocular, subtraiu-se da imagem com os mapas de atenção visual a imagem original do teste. Dessa maneira, obteve-se apenas o mapa de olhar de cada indivíduo (e.g ver Figura 1) para cada imagem (totalizando 40 imagens existentes no teste, por voluntário). Após binarizar o mapa de atenção visual encontrado (fixando o valor de limiar igual a 5 para tons de cinza entre 0 e 255), utilizou-se o método PCA + MLDA [Thomaz et al. 2006] para redução de dimensionalidade dos dados (considerando somente as componentes principais cujos autovalores são não-nulos), projeção das amostras no hiperplano discriminante de separação entre os grupos proficientes (quartil 4) e não-proficientes (quartis 1,2 e 3) e classificação pela distância euclidiana com relação à media de cada grupo para cálculo da taxa de acurácia.

3. Resultados

Analogamente ao resultado de [Bobak 2016] utilizando o CFMT+ [Russell et al. 2009], também foi encontrado aqui uma distribuição normal dos voluntários postos ao GFMT curto [Burton et al. 2010] em relação à frequência de acertos obtidos, como ilustrado na Figura 2.

Foi calculado o coeficiente de correlação de Pearson para analisar a relação entre o tempo e o rendimento dos voluntários submetidos ao GFMT [Burton et al. 2010]. Notou-se baixa correlação (valores absolutos menores que 0,7) nessa análise, o que indica que embora quanto mais rápido o voluntário for, melhor, isso não é expressivo estatisticamente. Em verdade, é possível observar que voluntários que obtiveram a mesma quantidade de acertos no teste tiveram coeficientes de Volke [Volke et al. 2002] muito

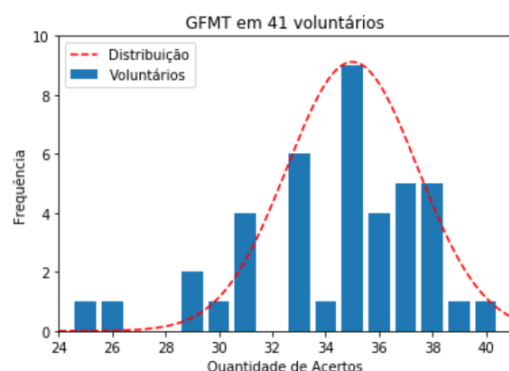


Figure 2. Dados obtidos durante aplicação de GFMT em participantes.

diferentes. Portanto, há tempos de execução muito variados (o tempo médio dos voluntários na execução do teste foi de 5 minutos e 49 segundos), fato observado por [Sergent and PONCET 1990] em seus pacientes com prosopagnosia, e aqui também em controles.

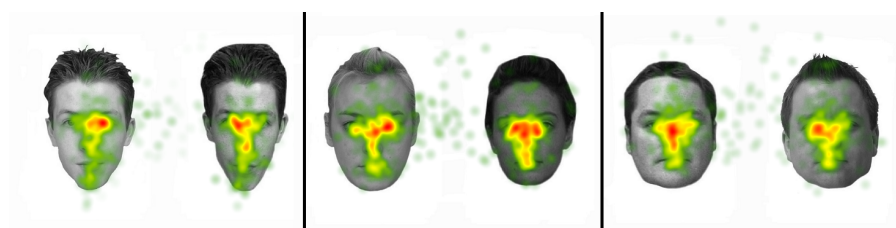


Figure 3. Estratégia ocular média entre 39 participantes submetidos ao GFMT. Quanto mais vermelho for o ponto no gráfico, mais a atenção do voluntário voltou-se para esse local.

A Tabela 1 mostra as correlações entre os fatores externos existentes nesse trabalho. Faz-se importante salientar a influência do sexo biológico nos resultados obtidos. Apesar da correlação de Pearson dessa influência externa ser baixa (menor que 0,7), ao analisar a estratégia ocular média entre os sexos, na Figura 5, nota-se, no público masculino, um leve deslocamento para a região central da face. Outro aspecto interessante é o cansaço físico, que, por mais que não possua correlação muito alta, estudos futuros poderiam observar mais atentamente esse efeito. Não houve voluntários fumantes e nem diagnosticados com prosopagnosia nesse experimento.

Por meio do equipamento de rastreamento de olhar, foi possível observar as características da estratégia ocular entre todos os participantes, por quartil e individualmente. Como mostra a Figura 3, o foco dos participantes, em média, foi detectado na região dos olhos, resultado observado por [Caldara et al. 2005] em seus voluntários controles e também relatado em [Chan et al. 2018]. Porém, ao observar o comportamento por quartis é possível visualizar que, quanto maior a proficiência do voluntário, mais a atenção volta-se para a região central do rosto (e.g a Figura 4 ilustra o comportamento ocular em uma questão do teste GFMT [Burton et al. 2010] com baixo índice de erro, já a Figura 6 ilustra o comportamento ocular em uma questão com alto índice de erro), o que demonstra que, como observado nos estudos de [Hsiao and Cottrell 2008] e [Peterson and Eckstein 2012], a atenção mais voltada ao nariz está provavelmente ligada

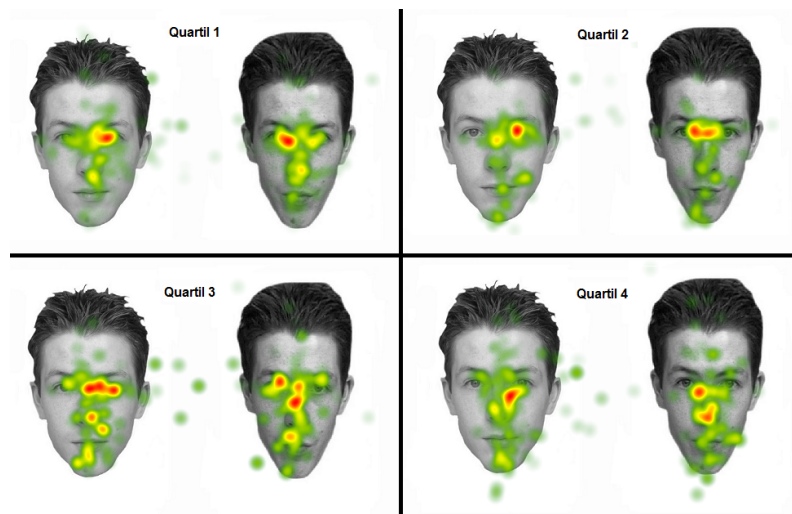


Figure 4. Estratégia ocular média por Quartis de proficiência em imagem com baixo índice de erro do GFMT [Burton et al. 2010].

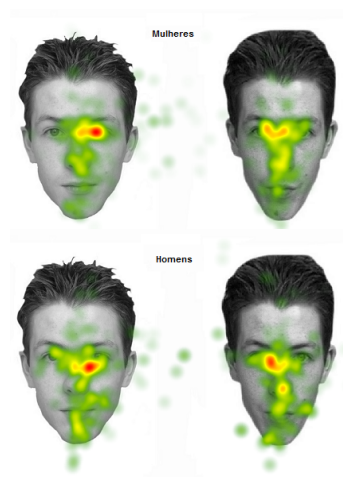


Figure 5. Influência do sexo biológico na estratégia de olhar média dos participantes mulheres e homens.

a alto desempenho de reconhecimento, e, de acordo com o presente estudo, no GFMT [Burton et al. 2010] também. A Figura 7 ilustra individualmente esse comportamento. Note as diferenças oculares do voluntário com a pior proficiência, medida pelo critério de Volke [Volke et al. 2002], do Quartil 1 (piores voluntários por quantidade de acerto) em contraste com o voluntário que obteve a maior proficiência, medida pelo critério de Volke [Volke et al. 2002]), do quartil 4 (melhores voluntários por quantidade de acertos). Esse resultado indica que o voluntário que acertou um número considerável de questões em pouco tempo, possivelmente, possui a estratégia ocular aproximada da ótima.

Na classificação de voluntários do quartil 4 com os demais observou-se, utilizando 10 iterações de validação cruzada com 90% dos dados utilizados para treino e 10% para inferência [Bengio and Grandvalet 2004], **94,8%** (DP = 2,9%) de acurácia na classificação de uma imagem com baixo índice de erros no experimento, ilustrado na Figura 8, e **91,47%** (DP=3,1%) de acurácia na classificação para uma imagem com alto

Table 1. Tabela de influências externas em relação aos acertos obtidos.

Influência	r de Pearson
<i>Idade</i>	0.069
<i>Sexo biologico</i>	0.315
<i>Escolaridade</i>	0.083
<i>Oculos</i>	−0.276
<i>Fome</i>	−0.079
<i>Cansaço físico</i>	−0.448
<i>Cansaço mental</i>	−0.234
<i>Medicamento</i>	−0.082
<i>Fumo</i>	—
<i>Prosopagnosia</i>	—

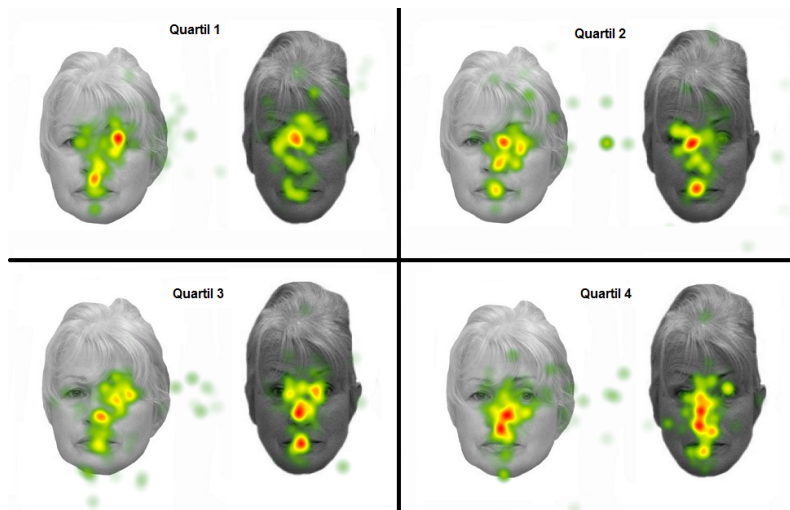


Figure 6. Estratégia ocular média por Quartis de proficiência em imagem com alto índice de erro do GFMT.

índice de erros no experimento, ilustrado na Figura 9.

4. Conclusão

Esse trabalho abordou o processamento holístico em imagens de faces, mostrando que os participantes, em média, e de acordo com a base analisada, observaram os olhos do indivíduo, porém, aqueles com rendimento superior no teste GFMT, tenderam a observar mais atentamente a região do nariz. Essas áreas da face, olhos e nariz, já foram evidenciadas como relevantes na literatura afim [Bobak 2016]. No entanto, para o processamento holístico avaliado, a estratégia ocular mostrou-se uma boa métrica para análise e classificação de proficiência no GFMT. Acredita-se que compreender as estratégias de quem reconhece muito bem pode contribuir no desenvolvimento de novos algoritmos capazes de reconhecer automaticamente faces e permitir o desenvolvimento de novos testes para encontrar super-reconhecedores entre nós. Trabalhos futuros poderão analisar o efeito da estratégia ocular sob influência de outros tipos de processamento facial, tais como configural, por componentes/características locais (featural), incluindo esses processamentos na percepção de emoções [Bombari et al. 2013] e no reconhecimento facial,

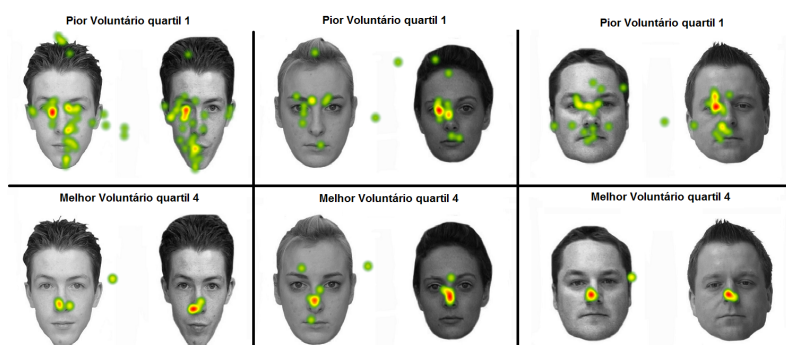


Figure 7. Mapa de movimentos oculares do melhor indivíduo do quartil 4 e do pior indivíduo do quartil 1, medidos pelo critério de Volke [Volke et al. 2002] postos ao GFMT.

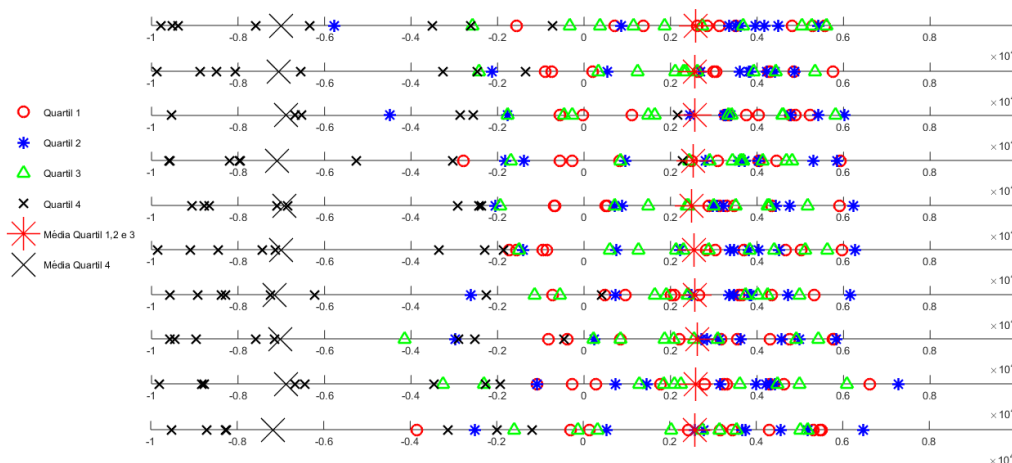


Figure 8. Eixo de discriminância obtido pelo MLDA em 10 iterações para questão com baixo índice de erros do GFMT .

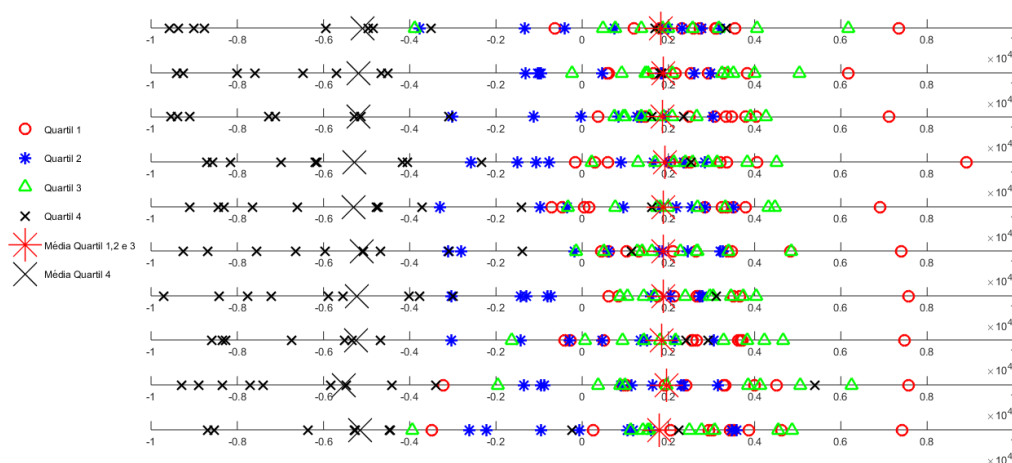


Figure 9. Eixo de discriminância obtido pelo MLDA em 10 iterações para questão com alto índice de erros do GFMT.

e também para classificação de sujeitos de fato super-reconhecedores e/ou diagnosticados com prosopagnosia adquirida ou congênita.

References

- Bate, S. and Cook, S. J. (2012). Covert recognition relies on affective valence in developmental prosopagnosia: Evidence from the skin conductance response. *Neuropsychology*, 26(5):670.
- Bate, S., Parris, B., Haslam, C., and Kay, J. (2010). Socio-emotional functioning and face recognition ability in the normal population. *Personality and Individual Differences*, 48(2):239–242.
- Bengio, Y. and Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 5(Sep):1089–1105.
- Bobak, A. K. (2016). *Theoretical and real-world applications of superior face recognition*. PhD thesis, Bournemouth University.
- Bodamer, J. (1947). Die prosop-agnosie. *Archiv für Psychiatrie und Nervenkrankheiten*, 179(1-2):6–53.
- Bombardi, D., Schmid, P. C., Schmid Mast, M., Birri, S., Mast, F. W., and Lobmaier, J. S. (2013). Emotion recognition: The role of featural and configural face information. *The Quarterly Journal of Experimental Psychology*, 66(12):2426–2442.
- Bruce, V. and Young, A. W. (2012). *Face perception*. Psychology Press.
- Burton, A. M., White, D., and McNeill, A. (2010). The glasgow face matching test. *Behavior Research Methods*, 42(1):286–291.
- Caldara, R., Schyns, P., Mayer, E., Smith, M. L., Gosselin, F., and Rossion, B. (2005). Does prosopagnosia take the eyes out of face representations? evidence for a defect in representing diagnostic facial information following brain damage. *Journal of cognitive neuroscience*, 17(10):1652–1666.
- Chan, C. Y., Chan, A. B., Lee, T. M., and Hsiao, J. H. (2018). Eye-movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic bulletin & review*, pages 1–8.
- Duchaine, B. and Nakayama, K. (2006). The cambridge face memory test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, 44(4):576–585.
- Gaur, R. P. and Jariwala, K. N. (2014). A survey on methods and models of eye tracking, head pose and gaze estimation. In *Journal of Emerging Technologies and Innovative Research*, volume 1. JETIR.
- Gobbini, M. I. and Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1):32–41.
- Grüter, T., Grüter, M., and Carbon, C.-C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *Journal of Neuropsychology*, 2(1):79–97.
- Hsiao, J. H.-w. and Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, 19(10):998–1006.
- Jones, R. and Tranel, D. (2001). Severe developmental prosopagnosia in a child with superior intellect. *Journal of Clinical and Experimental Neuropsychology*, 23(3):265–273.

- Peterson, M. F. and Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109(48):E3314–E3323.
- Robertson, D. J., Noyes, E., Dowsett, A. J., Jenkins, R., and Burton, A. M. (2016). Face recognition by metropolitan police super-recognisers. *PloS one*, 11(2):e0150036.
- Russell, R., Duchaine, B., and Nakayama, K. (2009). Super-recognizers: People with extraordinary face recognition ability. *Psychonomic bulletin & review*, 16(2):252–257.
- Sergent, J. and PONCET, M. (1990). From covert to overt recognition of faces in a prosopagnosic patient. *Brain*, 113(4):989–1004.
- Thomaz, C. E., Amaral, V., Giraldi, G. A., Gillies, D. F., and Rueckert, D. (2017). Is human face processing a feature-or pattern-based task? evidence using a unified computational method driven by eye movements. *arXiv preprint arXiv:1709.01182*.
- Thomaz, C. E., Kitani, E. C., and Gillies, D. F. (2006). A maximum uncertainty lda-based approach for limited sample size problems-with application to face recognition. *Journal of the Brazilian Computer Society*, 12(2):7–18.
- Volke, H.-J., Dettmar, P., Richter, P., Rudolf, M., and Buhss, U. (2002). On-coupling and off-coupling of neocortical areas in chess experts and novices as revealed by evoked eeg coherence measures and factor-based topological analysis—a pilot study. *Journal of Psychophysiology*, 16(1):23.
- White, D., Phillips, P. J., Hahn, C. A., Hill, M., and O’Toole, A. J. (2015). Perceptual expertise in forensic facial image comparison. *Proc. R. Soc. B*, 282(1814):20151292.