

## DETECÇÃO AUTOMÁTICA DO MANHÊS: ANÁLISE DA PROSÓDIA DE PAIS DE CRIANÇAS AUTISTAS

Ammar Mahdhaoui<sup>1</sup>, Mohamed Chetouani<sup>1</sup>, Erika Parlato-Oliveira<sup>2</sup>, Raquel S. Cassel<sup>3,6</sup>, Catherine Saint-Georges<sup>2,6</sup>, Marie Christine Laznik<sup>4</sup>, Fabio Apicella<sup>5</sup>, Filippo Muratori<sup>5</sup>, Sandra Maestro<sup>5</sup>, David Cohen<sup>3,6</sup>

ammar.mahdhaoui@upcm.fr; mohamed.chetouani@upmc.fr; eparlato@hotmail.com;  
raquelcassel@gmail.com; catherine.saint-georges@orange.fr; mclaznik@club-internet.fr;  
fapicella@inpe.unipi.it; filippo.muratori@inpe.unipi.it; david.cohen@psl.ap-hop-paris.fr;  
smaestro@inpe.unipi.it

**RESUMO:** O autismo é uma síndrome clínica bem definida após o segundo ano de vida, porém ainda são escassas as informações sobre o autismo nos dois primeiros anos de vida. O estudo de vídeos caseiros descreveu crianças com autismo durante o primeiro ano de vida, que não exibiam o padrão rígido de sintomas. Portanto, fatores desenvolvimentais e ambientais, além de fatores genéticos/biológicos parecem influenciar o aparecimento do autismo. Aqui nós descrevemos (1) uma hipótese que focaliza a possível implicação do empobrecimento do manhês durante a interação mãe/bebê como um possível co-fator, (2) a abordagem metodológica foi utilizada para desenvolver um algoritmo informatizado para detectar o manhês em vídeos caseiros; (3) o melhor desempenho de configuração do detector na extração do manhês a partir de seqüências de vídeos caseiros (precisão = 82%, em falante independente versus 87,5% em falante-dependente) deve ser utilizado para testar esta hipótese.

**PALAVRAS-CHAVE:** Manhês; Autismo; Detector automático de prosódia.

---

<sup>1</sup> Institut des Systèmes Intelligents et de Robotique, CNRS UMR 7222, Université Pierre et Marie Curie, Paris, France.

<sup>2</sup> Departamento de Fonoaudiologia, Universidade Federal de Minas Gerais – UFMG.

<sup>3</sup> Department of Child and Adolescent Psychiatry, AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Université Pierre et Marie Curie, Paris, France.

<sup>4</sup> Department of Child and Adolescent Psychiatry, Association Santé Mentale du 13ème, Paris, France.

<sup>5</sup> Scientific Institute Stella Maris, University of Pisa, Italy.

<sup>6</sup> Laboratoire Psychologie et Neurosciences Cognitives, CNRS UMR 8189, Paris, France.

## **INTRODUÇÃO**

O autismo é uma síndrome psiquiátrica grave, caracterizada pela presença de anormalidades nas interações sociais recíprocas, padrões anormais de comunicação e um repertório de comportamentos, interesses e atividades restrito, estereotipado e repetitivo. Embora seja uma síndrome clínica bem definida após o segundo ano de vida e, especialmente, depois do terceiro, são escassas as informações sobre autismo nos dois primeiros anos de vida (Short e Schopler, 1988; Stone et al., 1994; Sullivan et al., 1990). Vídeos caseiros (ou seja, filmes gravados pelos pais durante os primeiros anos de vida, antes do diagnóstico) e observação direta de bebês de alto risco, devido ao fato de serem irmãos de crianças autistas, são as duas fontes de informação mais importantes. Esses dois métodos têm permitido descrever crianças com transtorno autista (TA) durante o primeiro ano de vida. Em particular, as crianças podem olhar as pessoas, procurar a fonte sonora de vozes e manifestar interesse em comunicação, como crianças com desenvolvimento normal fazem. É de grande importância obter mais detalhes sobre estas competências sociais e em quais situações elas surgem, preferencialmente em crianças que estão em risco de desenvolver o autismo. Neste artigo, vamos nos concentrar em um estímulo verbal chamado manhês (um tipo especial de discurso que é dirigido às crianças) que tem demonstrado ser de importância crucial na aquisição da linguagem (Kuhl, 2004), e investigarmos se o seu curso durante a interação entre a mãe e a criança pode ser afetado pela criança autista.

### **1. AQUISIÇÃO DA LINGUAGEM E INTERAÇÃO SOCIAL**

Para explorar estas questões, vamos rever brevemente o que sabemos sobre a aquisição da língua, da interação social e das formas de integrar as dificuldades de crianças autistas em ambos os domínios durante o seu desenvolvimento. Estudos de aquisição da linguagem são numerosos demais para uma ampla revisão, no entanto, queremos destacar alguns pontos específicos de acordo com Kuhl (2007, 2004, 2003, 2000) e Goldstein et al. (2003). (1) Desde o nascimento, as crianças possuem a capacidade de discriminar todas as línguas no nível fonológico e reconhecer suas características prosódicas. (2) As crianças desenvolvem uma estratégia de aprendizagem baseada na entrada de sinais e de características lingüísticas e exploram as propriedades estatísticas da língua, levando à “aprendizagem probabilística” (Milgram e Atlan, 1983). (3) A experiência lingüística compromete o sistema perceptivo no nível neural, aumentando a percepção de fala da língua materna e diminuindo a

ReVEL, v. 8, n. 15, 2010. ISSN 1678-8931 149

percepção de fala em língua estrangeira (Kuhl, 2000). No entanto, a simples exposição não explica a aquisição de línguas: tanto na produção quanto na percepção da fala, a presença de um ser humano interagindo com uma criança tem forte influência sobre a aquisição (Goldstein et al., 2003).

Independentemente da influência dos fatores neurobiológicos e genéticos sobre o autismo (Cohen et al., 2005), deve-se considerar que a sobrevivência e o desenvolvimento de bebês dependem da interação social com um cuidador, que atende às necessidades do bebê. A qualidade da interação social depende de um processo recíproco, um diálogo ativo entre cuidador e bebê com base em competências precoces do bebê e da capacidade do cuidador para se adaptar. Micro-análises de registros visuais analisando os ritmos e os padrões de interação entre mãe e bebê, face a face, no período neonatal (Brazelton et al., 1975) e na comunicação precoce (Condon e Sander, 1974; Stern et al. 1975) enfatizaram a importância da sincronia e co-modalidade destas interações iniciais para o desenvolvimento do bebê.

## **2. CORRELAÇÃO ENTRE AQUISIÇÃO PRECOCE DE LÍNGUAS E INTERAÇÃO SOCIAL**

Curiosamente, pesquisadores de aquisição da linguagem e de interações sociais precoces têm encontrado uma peculiaridade importante, que afeta tanto a linguagem quanto o desenvolvimento social das crianças: a maneira como os adultos falam às crianças. Um tipo especial de discurso, que é direcionado às crianças, muitas vezes chamado de *manhês*, caracterizado por frequência mais alta, ritmo mais lento, e contornos de entonação exagerados (Fernald, 1985; Grieser e Kuhl, 1988, Mehler et Dupoux, 1995), parece desempenhar um papel importante tanto na interação social quanto no desenvolvimento da linguagem. Estudos têm revelado que esta prosódia particular pode ser responsável por atrair a atenção do bebê, transmitindo afeto emocional e informação fonológica específicas da língua (Karson, 1985; Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1989; Fernald, 1985; Fernald e Kuhl, 1987). (1) Quando dada uma escolha, as crianças demonstram preferência para o discurso materno em *manhês* em oposição à fala dirigida ao adulto (Pegg et al., 1992; e Aslin Cooper, 1990; Werker e McLeod, 1989; Fernald, 1985; e Glenn Cunningham, 1983). Esta prosódia particular favorece o interesse e possibilita manter a atenção do bebê. (2) O *manhês* contém particularmente bons exemplares fonéticos - os sons são mais claros e mais distintos entre si (devido por exemplo à hiperarticulação da vogal) -, quando comparado ao discurso dirigido ao adulto (Kuhl et al., 1997; Burnham et al., 2002).

Além disso, as reações do bebê amplificam os contornos das curvas prosódicas na voz da mãe (Burnham et al., 2002). Portanto, o manhês depende da qualidade da resposta dos bebês, o que sugere que as crianças estão ativamente envolvidas nas interações precoces (Braarud et al., 2008).

### **3. É POSSÍVEL IMPLICAR O EMPOBRECIMENTO DO MANHÊS NA PATOGÊNESE DO AUTISMO?**

Em nossa opinião, as deficiências conhecidas dos autistas em linguagem, cognição e desenvolvimento social, bem como as tendências para a auto-absorção, perseverança e auto-estimulação (Volkmar e Pauls, 2003), podem ter efeitos subjacentes nas dificuldades primárias na capacidade de participar em interações envolvendo sinais emocionais, gestos motores e atos comunicativos dirigidos aos outros. Essa visão também é apoiada pelo testemunho de indivíduos adultos com autismo (Chamak et al., 2008). Se a aprendizagem depende de interesse social entre as pessoas e dos sinais que elas produzem, as crianças com autismo, que apresentam menor interesse social, podem ter uma desvantagem cumulativa na aquisição da linguagem. Sua resposta escassa aos pais pode empobrecer o engajamento parental e também a produção do manhês, no decorrer do tempo. Como consequência, esse empobrecimento reforçará o retraimento social e o atraso na aquisição da linguagem. Em um estudo exploratório, observou-se seqüências de filmes caseiros, em que um bebê, que será posteriormente diagnosticado como autista, demonstra interesse quando o pai emprega uma expressão vocal em manhês. Durante essas interações, bebês e crianças com autismo apresentam atenção social focal: seus rostos se iluminam, inesperadas habilidades interativas aparecem, bem como verdadeiras protoconversações (Laznik et al., 2005). No entanto, não foi possível encontrar nenhum estudo comparativo sobre o interesse do bebê ao manhês versus discurso dirigido a adulto, em crianças com autismo. Mesmo sabendo que as crianças autistas podem processar alguns aspectos da voz humana (Groen et al. 2008), no entanto, estudos têm demonstrado que sujeitos autistas não apresentam ativação cortical específica em resposta às vozes humanas (Zilbovicius et al., 2006 ). Além disso, aos 5 anos de idade, as crianças autistas não demonstram a preferência prevista para o discurso materno, quando administrado como uma escolha com o barulho de vozes sobrepostas (Klin, 1991). Da mesma forma, entre 32 e 52 meses de idade, crianças com Transtorno do Espectro do Autismo têm preferência significativa para escutar sinais eletrônicos analógicos de manhês em comparação com sinais do manhês (Kuhl et al., 2005).

Considerando que filmes caseiros oferecem uma oportunidade única de acompanhar o desenvolvimento do bebê e da criança em interações com a mãe, nós planejamos usar uma abordagem multidisciplinar, com psiquiatria infantil, psicologia do desenvolvimento, psicolinguística e engenharia acústica para testar as seguintes questões de investigação: (1) As crianças, que mais tarde são diagnosticadas como autistas, inicialmente estariam equipadas para responder especificamente ao manhês? Em outras palavras, essa competência é estável e correlacionada a interações positivas? (2) Como a amplificação do manhês é bidirecional, a quantidade das horas de produção de manhês pelos pais, nestes casos, decresce continuamente?

#### **4. OBJETIVOS**

Neste trabalho, nós descrevemos (1) um algoritmo automático para detectar o manhês, (2) as suas características na extração do manhês a partir de seqüências de vídeos caseiros. Dado que a micro análise de vídeos caseiros é muito trabalhosa e demorada, consideramos a produção deste algoritmo como o primeiro passo do nosso projeto de pesquisa multidisciplinar. Na verdade, mesmo se o manhês estivesse claramente descrito em termos de propriedades acústicas, a modelagem e a detecção seria difícil, como é o caso com a maioria dos discursos de emoção, porque a caracterização da fala espontânea e afetiva ainda é uma questão, já que vários parâmetros têm sido propostos na literatura (Schuller et al., 2007). Como ponto de partida sobre as propriedades acústicas do manhês, nós caracterizamos as interações verbais a partir da extração de características supra-segmentais (prosódicas). Dado que os filmes caseiros não são registrados por profissionais, muitas vezes eles contêm condições adversas (por exemplo, no que diz respeito ao ruído, à câmera, ou aos microfones), mesmo assim, a segmentação acústica de filmes caseiros mostra que as características segmentares desempenham um papel importante em termos de robustez (Schuller et al, 2007). Os enunciados são caracterizados por características supra-segmentares (por exemplo, as estatísticas com relação à frequência fundamental, de energia e de duração). Estas características, juntamente com os rótulos conhecidos de um treinamento conjunto de expressões emocionais, são fornecidas como entrada para um algoritmo de aprendizagem da máquina. Aqui, nós comparamos o desempenho de dois algoritmos de aprendizagem, GMM (Gaussian Mixture Model) e k-nn (k-mais próximos), sobre as características segmentares e não segmentares isoladamente, usando o método de análise ROC (receiver operating characteristic). Para melhorar a detecção, também investigou-se várias combinações ou

ReVEL, v. 8, n. 15, 2010. ISSN 1678-8931 152

esquemas de fusão (segmentar /supra-segmentares; MGM / k-nn) para selecionar o que é melhor em comparação com a segmentação manual das seqüências de manhês em vídeos caseiros.

## **5. MÉTODO**

### **5.1 BANCO DE DADOS**

O corpus de fala utilizado neste experimento é composto por um conjunto de interações naturais e espontâneas. Este corpus contém expressões de comunicação transmitidas por pais de crianças em fase pré-verbal. O corpus é composto por gravações de mães e pais italianos com seus bebês. Decidimos centrar a análise em filmes caseiros, pois ela permite um estudo longitudinal e dá informações sobre o comportamento precoce das crianças autistas muito antes do diagnóstico feito por médicos. Todas as seqüências foram extraídas do banco de dados de vídeos caseiros de Pisa, que inclui filmes caseiros dos primeiros 18 meses de vida de três grupos de crianças de 4-7 anos (Maestro et al. 2005b). As crianças são pareadas por sexo e condição sócio-econômica. O primeiro grupo foi composto por 15 crianças (M/F:05/10) com diagnóstico de autismo com base no Autismo Diagnostic Interview-Revised (ADI-R) (Rutter, 2003). O segundo grupo foi composto por 12 crianças (M/F:05/07) com diagnóstico de Retardo Mental (RM) com características não-autísticas. Crianças com RM ou AD secundário a síndromes neuropsiquiátricas (por exemplo, do X-Frágil, Rett, ou Síndrome de Down) ou evidentes déficits neurológicos ou ainda portadoras de anomalias físicas foram excluídas. Os participantes do grupo AD e do grupo MR foram submetidos à Escala de desenvolvimento de inteligência Griffiths Mental e Escala Wechsler, a fim de determinar o funcionamento intelectual. Tanto o grupo AD (QI médio: 59,26; dp: 8,49) quanto o grupo MR (QI médio: 56,82; dp: 8,16), obtiveram escore de QI abaixo de 70. Um terceiro grupo foi composto por 15 crianças de desenvolvimento típico (DT) (M/F:06/09) recrutada em uma creche local.

O grande tamanho do corpus, no entanto, é inconveniente para a análise manual das características prosódicas. Para o desenvolvimento do primeiro algoritmo, nós nos concentramos em um vídeo caseiro de qualidade acústica superior, que totalizaram 3 horas do primeiro ano de vida de uma criança, posteriormente diagnosticada como autista (aos 6 anos: ADI-R pontuação comunicação = 9; ADI- R pontuação interesse social = 12; ADI-R repetidas pontuação interesse = 5; CARS pontuação total = 36; QI = 71). As interações verbais da mãe

ReVEL, v. 8, n. 15, 2010. ISSN 1678-8931 **153**

da criança foram cuidadosamente anotadas por duas psicolingüistas, às cegas, em duas categorias: fala dirigida à criança (manhês) e fala dirigida a adultos (normal) ( $\kappa$  de Cohen = [0,82 95% CI: 0,75-0,90], CI para Intervalo de Confiança). A partir desta anotação manual, nós extraímos 100 declarações para cada categoria. Os enunciados têm tipicamente entre 0,5s e 4s de duração. Para realizar uma segunda mostra característica do estudo, coletamos um conjunto independente de declarações (sendo 50 de manhês e 50 de fala normal), extraídas de dez filmes caseiros selecionados aleatoriamente (5 do grupo AD e 5 do grupo TD).

## 5.2 DESCRIÇÃO DO SISTEMA

A segmentação automática dos sons da fala requer várias etapas, que são realizadas seqüencialmente: extração das características, classificação e decisão da fusão. Essas etapas podem ser divididas em duas fases principais de processamento, ou seja, a fase de treinamento e a fase de teste ou, fase de classificação. A Figura 1 mostra a seguir, de forma esquemática, o que é descrito com maiores detalhes. O primeiro passo, ou seja, a extração de características, mostrado na Figura 1 é necessário tanto no treinamento quanto no teste. A segunda etapa, denominada classificação, visa classificar todos os enunciados. A última etapa, que é chamada de fusão/decisão, é necessária para a classificação e combina os diferentes fluxos para formar uma saída comum.

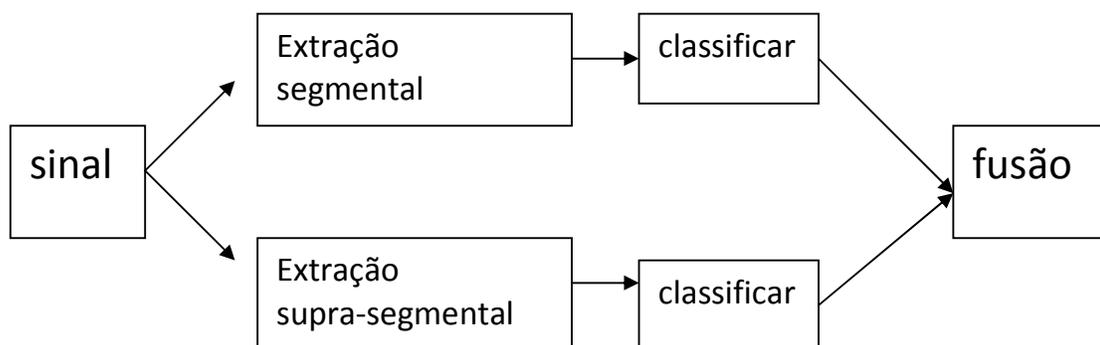


Figura 1: Sistema de classificação do manhês usado no algoritmo.

## 5.3 EXTRAÇÃO

Após a segmentação da base de dados em expressões, o próximo passo é extrair características, resultando em uma representação do sinal de voz como entrada para o sistema

de classificação do manhês. Neste trabalho avaliou-se duas abordagens, chamadas, respectivamente, características segmentais e supra-segmentares. A primeira é caracterizada pela Coeficiente de frequência Mel cepstro (MFCC), enquanto a segunda é caracterizada por medidas estatísticas tanto da frequência fundamental (F0) quanto da energia a curto prazo. Uma janela de 20ms é usada, e as sobreposições entre os quadros adjacentes é 02/01. Um vetor de parâmetros de ordem 16 foi calculado. As características supra-segmentares são caracterizadas por três estatísticas (média, variância e intervalo) em ambos os F0 e curto prazo de energia, resultando em um vetor de 6-dimensional. Note-se que a duração dos eventos acústicos não é diretamente caracterizada como um recurso, mas é considerada durante o processo de classificação, por um fator de ponderação. Os vetores de características são normalizados (média zero, uma unidade de desvio padrão).

#### 5.4 CLASSIFICAÇÃO

Neste estudo, dois classificadores diferentes: o K-nn e Modelo Misto de Gaussian (GMM) foram usados. O K-nn é um classificador frequentemente usado em reconhecimento de padrões. GMM é um modelo estatístico de sinais formados por seqüências de vetores de características que são representativos do sinal de entrada (Reynolds, 1995). (1) Probabilidade Estimada *a posteriori*: o GMM é adotado para representar a distribuição dos recursos. Partindo do pressuposto de que a seqüência característica do vetor  $x = (x_1, x_2, \dots, x_n)$  é uma seqüência de distribuição independente, a estimativa de distribuição do vetor d-dimensional característica  $x$  é uma soma ponderada das densidades gaussianas do componente  $M$   $g(\mu, \Sigma)$ , cada um parametrizado por um vetor de média e covariância  $\mu_i$   $\Sigma_i$  matriz, a densidade da mistura para o modelo de  $C_m$  é definido como:

$$P(x | C_m) = \sum_{i=1}^M w_i g(\mu_i, \Sigma_i)(x) \quad (1)$$

Cada componente da densidade é uma função d-variável gaussiana:

$$g_{(\mu, \Sigma)}(x) = \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Sigma)}} e^{-1/2(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2)$$

A mistura dos pesos  $w_i$  satisfaz a restrição seguinte:

$$\sum_{i=1}^M w_i = 1 \quad (3)$$

O vetor de características  $x$  é então modelado, *a posteriori*, pela seguinte probabilidade:

$$P_{gmm}(C_m | x) = \frac{p(x | C_m) p(C_m)}{p(x)} \quad (4)$$

Onde  $P(C_m)$  é a probabilidade prévia para a classe  $C_m$ , supomos igual probabilidade prévia. Nós usamos o algoritmo de maximização da expectativa (EM) para a mistura da obtenção da máxima verossimilhança. O classificador K-nn é uma técnica não-paramétrica que classifica o vetor de entrada com o rótulo da maioria k-nn (protótipos) (Duda et al., 2000). A fim de manter um quadro comum com o classificador estatístico (GMM), estimamos a probabilidade *a posteriori* de que um vetor dado de características  $x$  pertence à classe  $C_m$  usando a estimativa K-nn (Duda et al., 2000):

$$P_{knn}(x | C_m) = \frac{k_m}{k} \quad (5)$$

onde  $k_m$  denota o número de protótipos que pertencem à classe  $C_m$  entre os k-nn.

(2) Segmentação baseada nos traços segmentares e supra-segmentares: características Segmental (ie, MFCC) são extraídos de todos os quadros de um enunciado  $U_x$  de forma independente para as partes manifestas ou não. Note-se que a natureza dos segmentos também pode ser explorada (vogais/consoantes) (Ringeval e Chetouani, 2008). Probabilidades *a posteriori* são, então, estimadas por ambos classificadores: MGM e K-nn e são chamadas, respectivamente, de  $P_{gmm,seg}(C_m | U_x)$  e de  $P_{knn,seg}(C_m | U_x)$ . A classificação das funções supra-segmentares segue a abordagem baseada em segmentos (SBA) (Shami e Verhelst, 2007). Um enunciado  $U_x$  é segmentado em  $N$  segmentos sonoros ( $F_{xi}$ ), obtidos por extração de  $F_0$  (veja acima). Estimativa local de probabilidades *a posteriori* é efetuada para cada segmento. A classificação combina as estimativas de emissão de  $N$  local.

$$P(C_m | U_x) = \sum_{x_i=1}^N P(C_m | F_{x_i}) \times length(F_{x_i}) \quad (6)$$

A duração dos segmentos é apresentada como o peso das probabilidades *a posteriori*: a importância do segmento de voz (comprimento ( $F_{xi}$ )). A estimativa é também realizada

pelos dois classificadores, resultando em caracterizações supra-segmentares:  $P_{gmm, supra}(C_m|U_x)$  e  $P_{K-nn, supra}(C_m|U_x)$ .

## 5.5 FUSÃO

As características segmentares e supra-segmentares fornecem informações temporais diferentes, e combiná-las deve melhorar a precisão do detector. Muitas técnicas de decisão podem ser empregadas (Kuncheva, 2004), mas nós investigamos uma simples soma ponderada das probabilidades dos classificadores diferentes:

$$C_l = \lambda \cdot \log(P_{seg}(C_m|U_x)) + (1 - \lambda) \cdot \log(P_{supra}(C_m|U_x)) \quad (7)$$

com  $l = 1$  (manhês) ou  $2$  (discurso dirigido a adulto). O coeficiente de ponderação é representado por  $\lambda$ . Para o classificador MGM, as probabilidades podem ser facilmente calculadas a partir das probabilidades *a posteriori* ( $P_{gmm, seg}(C_m|U_x)$ ,  $P_{gmm, supra}(C_m|U_x)$ ) (Reynolds, 1995). No entanto, a estimativa de K-nn pode produzir uma probabilidade nula *a posteriori* incompatível com o cálculo da probabilidade. Usamos uma solução recentemente testada por Kim et al. (2007), em que a probabilidade *a posteriori* é usada em vez da probabilidade log do K-nn:

$$C_l = \lambda \cdot \log(e^{P_{km, seg}(C_m|U_x)}) + (1 - \lambda) \cdot \log(P_{gmm, supra}(C_m|U_x)) \quad (8)$$

Combinação 1	$P_{K-nn, seg}$	$P_{K-nn, supra}$
Combinação 2	$P_{gmm, seg}$	$P_{gmm, supra}$
Combinação 3	$P_{K-nn, seg}$	$P_{gmm, supra}$
Combinação 4	$P_{gmm, seg}$	$P_{K-nn, supra}$
Combinação 5	$P_{gmm, seg}$	$P_{K-nn, seg}$
Combinação 6	$P_{gmm, supra}$	$P_{K-nn, supra}$

**Tabela 1:** Tabela de combinação de classificadores (GMM / k-nn) x características (segmental / suprasegmental). GMM = Modelo Misto de Gaussian; k-nn = k-nearest neighbours; seg = segmental; supra = suprasegmental

Para o classificador k-nn usamos a equação 7, enquanto que para o MGM a probabilidade foi convencionalmente calculada. Nós investigamos então as combinações híbridas.

## 6. RESULTADOS

### 6.1 CLASSIFICADOR DE CONFIGURAÇÃO

Inicialmente, para encontrar a melhor estrutura de nossa classificação, tivemos que ajustar os parâmetros: o número de gaussianas (M) para o classificador GMM e o número de elementos (K) para o classificador K-nn. Procuramos a configuração ideal em termos de precisão. A Tabela 2 mostra a melhor configuração da MGM e K-nn com características segmentais e supra-segmentares e revela ainda que o classificador GMM foi treinado com recursos prosódicos que superaram os classificadores, em termos de precisão.

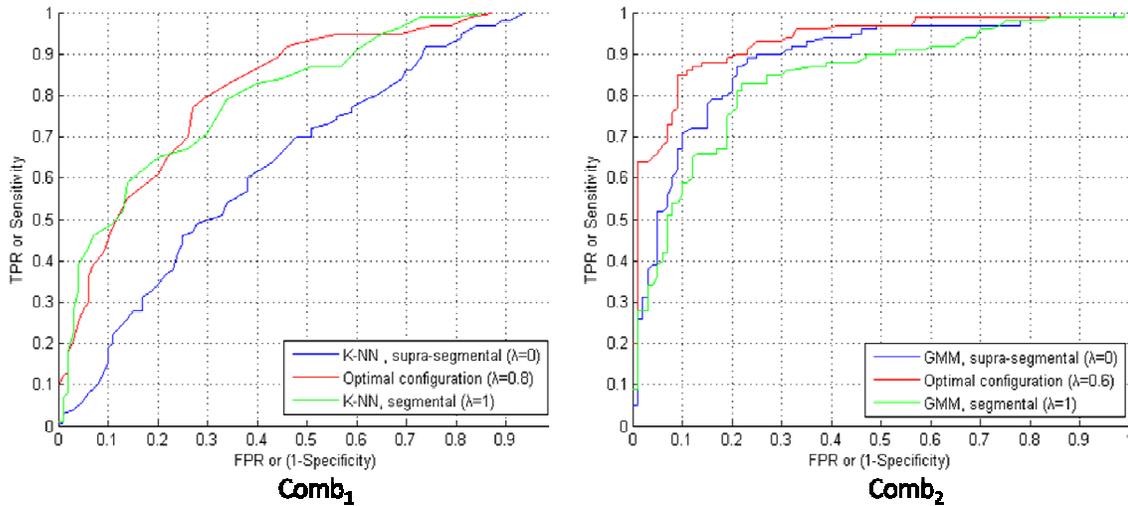
	Segmental	Supra-Segmental
k-nn	72.5% (k=11)	61% (k=7)
GMM	78%(M=15)	82%(M=16)

**Tabela 2:** Acuracia da melhor configuração para os classificadores GMM e k-nn de acordo com as características segmental ou suprasedgmental. GMM = Modelo Misto de Gaussiano; k-nn = k; M = número de Gaussianos para GMM; K = número de elementos para k-nn.

### 6.2 FUSÃO DE MELHOR SISTEMA

Neste estudo, o desempenho do classificador foi citado com sensibilidade de classe, preditivas e acuracia geral, e os classificadores ótimos foram determinados empregando gráficos ROC para mostrar o trade-off entre o sucesso e as taxas de falsos positivos. A curva ROC representa o equilíbrio entre a taxa de falsa aceitação (FAR) e a taxa de falsa rejeição como o limiar do classificador de saída que é variável. Duas medidas quantitativas de verificação do desempenho, a taxa de erro igual (EER) e a área sob a curva ROC (AUC) foram calculadas. Todos os cálculos foram realizados com Matlab (versão 6). Para obter melhores dados de configuração, os resultados foram administrados com intervalos de confiança de 95%, que foram estimados utilizando o Método de Cornuéjols et al. (2002). Deve-se notar que, embora EER represente o desempenho de um classificador em apenas um limiar de funcionamento, a AUC representa o desempenho global do classificador em todo o intervalo de limiares. Assim, foram empregados AUC e não EER para comparar o desempenho de verificação de dois classificadores e suas combinações. No entanto, o resultado mostrado na tabela 2 motivou uma investigação da fusão de ambas as características e classificadores após a

abordagem estatística descrita na seção anterior. A melhoria da combinação de características e classificadores é conhecida por ser eficiente (Kuncheva, 2004). No entanto, devemos ter cuidado, porque a fusão das melhores configurações nem sempre dá melhores resultados, a eficácia dependerá da natureza dos erros produzidos pelos classificadores (independente versus dependentes) (Kuncheva, 2004). A Tabela 1 e a seção anterior mostram que 6 diferentes regimes de fusão poderiam ser investigados (Comb1 para Comb6). Para cada um deles, nós aperfeiçoamos parâmetros classificadores (o valor de K para K-nn e o valor de M para GMM, respectivamente) e o coeficiente  $\lambda$  (eq. 6) para os parâmetros da fusão. Na figura 2, podemos ver que para o classificador K-nn, as melhores pontuações (0.8113/0.812) foram obtidas com uma contribuição importante dos recursos segmentares ( $\lambda = 0,8$ ), que está de acordo com os resultados obtidos sem a fusão (tabela 2). Os melhores resultados GMM (0.932/0.932) são obtidos com um fator de ponderação igual a 0,6, revelando um equilíbrio entre as duas características. A Tabela 3 resume os melhores resultados em termos de precisão, bem como o valor preditivo positivo (VPP) e o valor preditivo negativo (VPN) para cada classificador de fusão (parte superior) e de classificação cruzada de fusão (parte inferior). A tabela mostra que diferentes combinações de características segmentais e supra-segmentais podem atingir precisões similares com PPVs diferentes. No que diz respeito ao método de K-nn, as melhores pontuações (Precisão = 74%, VPP = 77,91%) foram obtidas com uma importante contribuição das características segmentais ( $\lambda = 0,9$ ). Os melhores resultados GMM (Precisão = 87,5%, VPP = 88,47%) foram obtidos com um fator de ponderação igual a 0,4 (ou seja, um equilíbrio entre as duas características). Também foram investigadas a fusão de classificação cruzada (Tabela 3, abaixo). A precisão deste método pode ser menor do que para os classificadores individuais (tabela 2). No entanto, melhorias significativas (90,7%) no PPV foram alcançadas para a Combinação 3: Pknn; seg x Pgmm; supra combinação ( $\lambda = 0,6$ ). Este resultado revela a importância da avaliação métricas para a fusão, que é dependente da tarefa neste estudo (detecção do manhês).



**Figura 2:** Curvas de ROC para Comb1 e Comb2

Combinação 1 =  $P_{K-nn,seg} \times P_{K-nn,supra}$  ; Combinação 2 =  $P_{gmm,seg} \times P_{gmm,supra}$  ;  $\lambda$  = coeficiente de peso usado na equação de fusão para cada combinação.

<i>USANDO APENAS UM CLASSIFICADOR</i>					
		Acuracia	PPV	NPV	$\Lambda$
Comb1 <i>K-nn, seg x k-nn,supra</i>	K-nn,seg (k=11)	74%	77.91%	71.05%	0.9
	K-nn,supra (k=1)				
	K-nn, seg(k=7) K-nn, supra (k=11)	72.5%	74.42%	71.84%	0.8
Comb2 <i>GMM, seg x GMM,supra</i>	GMM, seg (M=12)	87.5%	88.47%	86.41%	0.4
	GMM, supra(M=15)				
	GMM, seg (M=16)	86.5%	84.85%	88.42%	0.5
	GMM, supra(M=16)				
<i>COMBINANDO K-NN EGMM</i>					
Comb3 <i>K-nn, seg x GMM, supra</i>	K-nn, seg (k=11)	85.5%	89.41%	83.81%	0.7
	GMM, supra(M=12)				
	K-nn, seg(k=5) GMM, supra(M=12)	85%	90.7%	82.69%	0.6
Comb4 <i>GMM, seg x k-nn, supra</i>	GMM, seg(M=15)	80.5%	79.61%	81.44%	0.2
	K-nn, supra(k=11)				
	GMM, seg (M=16) K-nn, supra(k=1)	79.5%	77.57%	81.72	0.3

**Tabela 3:** Acuracia, valor de predição positiva (PPV), e valor de predição negativa (NPV) da fusão de configurações para classificadores de parâmetros segmental e supra-segmental. PPV (Valor de Predição positiva) = TruePositive / (TruePositive+FalsePositive). PPV avalia a proporção de detecção correta de enunciados em manhês entre todos os enunciados classificados como manhês. PPV pode ser visto como a reclassificação das predições positivas induzidas pelo classificador. NPV (Valor preditivo negativo) possibilita informação similar sobre a detecção da fala dirigida a adulto.

Em suma, a melhor configuração de fusão utiliza apenas o classificador MGM tanto para as características segmentares quanto para supra-segmentares (M=12 e M=15, respectivamente,  $\lambda = 0,4$ ). O desempenho desta fusão corresponde à combinação 2 na Tabela 3, e foram os seguintes: Exatidão = 87,5% (IC 95%: 82,91% -92,08%), VPP = 88,47% (IC 95%: 83,03% -95,18%); NPV = 86,41% (95% IC: 79,4% para -92,88%). Também foi ReVEL, v. 8, n. 15, 2010.

investigado o desempenho do detector de fusão com esses parâmetros para detectar manhês vs fala dirigida a adulto em um segundo conjunto de afirmações extraídas de dez vídeos caseiros aleatoriamente selecionados, com 12 falantes independentes. As performances em condições de falantes independentes foram as seguintes: Precisão = 82% (IC 95%: 73,87% -89,58%), VPP = 86,36% (IC 95%: 66,52% -89,48%); NPV = 77,55% (IC 95% : 76,73% -96,6%).

## **7. DISCUSSÃO**

### **7.1. VALIDADE DO DETECTOR DE MANHÊS**

Nosso objetivo foi desenvolver um detector de manhês investigando diferentes características e classificadores. Utilizando técnicas de classificação que são frequentemente utilizadas para reconhecimento de fala e de falante (GMM e K-nn). Nós desenvolvemos um sistema de detecção de manhês e testamos no modo dependente do falante. A fusão de características e classificadores também foi investigada. Obtivemos resultados a partir dos quais podemos inferir conclusões interessantes. Em primeiro lugar, nossos resultados mostram que as características segmentares sozinhas contêm informação útil para a discriminação entre manhês e discurso direto a adultos, da partir de suas características supra-segmentares. Assim, podemos concluir que as características segmentares podem ser usadas de forma exclusiva. No entanto, de acordo com nossos resultados de detecção, características prosódicas também são muito promissoras. Com base nas duas conclusões anteriores, nós combinamos classificadores que utilizam recursos segmentares com os classificadores que utilizam recursos supra-segmentares e descobrimos que esta combinação melhora o desempenho do nosso detector de manhês consideravelmente.

Assim, podemos concluir que um classificador baseado em características segmentares só pode ser usado para discriminar entre manhês e fala dirigida a adulto, mas uma melhora significativa pode ser alcançada na maioria dos casos, quando este classificador é fundido com um outro que é baseado em características prosódicas . Além disso, os escores de fusão do K-nn e MGM também é muito produtiva.

### **7.2. LIMITAÇÕES E PONTOS FORTES**

Para nossos experimentos de classificação de manhês, usamos apenas trechos de fala que já foram segmentados (com base na transcrição humana). Em outras palavras, a detecção

ReVEL, v. 8, n. 15, 2010. ISSN 1678-8931 161

de início e deslocamento de manhês não foi investigada neste estudo, mas pode ser abordada em um estudo futuro. Detecção do início e deslocamento de manhês (segmentação do manhês) pode ser vista como um problema distinto que dá origem a outras questões interessantes, tais como a forma de definir o início e o fim do manhês e quais tipos de medidas de avaliação utilizar. Estes são problemas que normalmente podem ser tratados dentro do modelo de Hidden Markov.

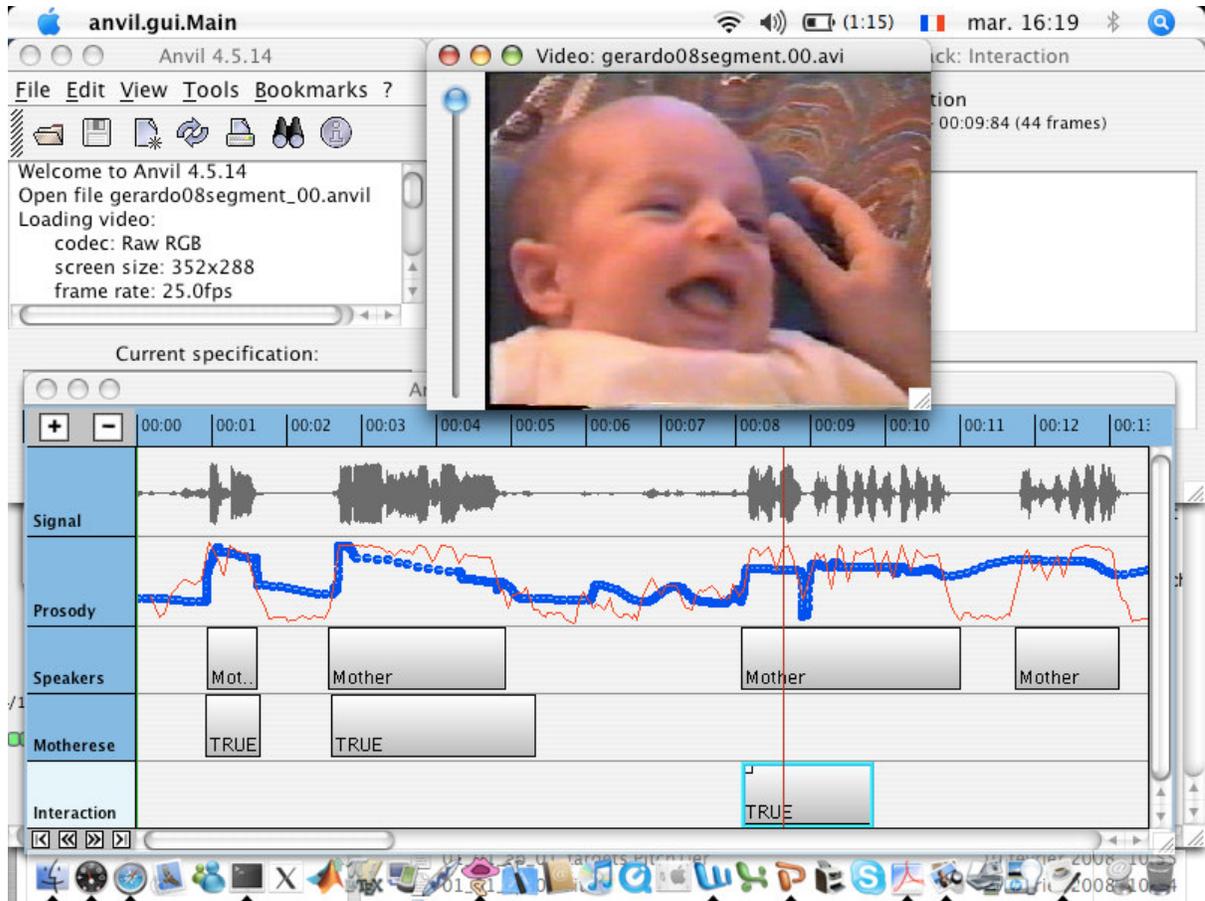
Na sua melhor configuração, o detector tinha valor preditivo atingindo 90%, esse nível de predição é adequado para estudos de vídeos caseiros. Quando nós exploramos o desempenho do detector de manhês em sequências cegamente validadas por dois psicolingüistas (falante independente), o desempenho do detector permaneceu muito bom (precisão = 82%). À medida que a hipótese do detector for independente da língua, no entanto, ainda é preciso explorar o desempenho do detector de manhês em sequências extraídas aleatoriamente de diferentes línguas. Como se observa, embora o manhês seja um sinal altamente variável e complexo (Fernald e Kuhl, 1987), seria interessante investigar até que ponto ele depende do falante e da cultura, entre outras variáveis. Finalmente, o objetivo deste estudo foi desenvolver um detector de manhês para permitir a classificação da emoção. Nosso plano de reconhecimento de emoção é baseado no fato de que a emoção é expressa em discurso sonoro a partir de eventos ou recursos paralingüísticos, que podem ser vistos como "recursos de emoção" em particular. A informação visual poderia então ajudar a melhorar a detecção da emoção. Nosso plano é, portanto, classificar a emoção através da detecção de eventos sonoros paralingüísticos. Neste estudo, nós desenvolvemos um detector de manhês para fornecer um primeiro passo neste plano para um classificador de emoção no futuro.

### **7.3. APLICAÇÕES POSSÍVEIS PARA O CAMPO DO AUTISMO**

Apesar das limitações acima, estamos agora explorando sistematicamente em vídeos caseiros da base de dados de Pisa (Maestro et al., 2005b), os discursos naturais de manhês e a interação mãe-bebê e sua co-ocorrência. O padrão das interações pais-bebê tem sido amplamente explorado em recém-nascidos posteriormente diagnosticados com autismo (para revisão, ver Saint-Georges et al.; Palomo et al., 2006). No entanto, o manhês positivo e as interações de co-ocorrência ainda não foram investigadas. Isso será testado em sequências aleatórias extraídas do banco de dados usando comparações entre os grupos pareados por idade e sexo: o autismo versus desenvolvimento típico; autismo versus crianças com

deficiência intelectual, autismo de início precoce versus autismo de início tardio. Estas análises são possíveis graças à mostra formada por vários casos, pareados por idade e sexo com controles normais e com atraso no desenvolvimento (Maestro et al. 2005b). Nós supomos que o manhês deve ser correlacionado com interações positivas, tanto em autistas como em não-autistas, mas em menor grau no autismo.

A segunda análise se concentrará em uma investigação longitudinal do manhês usando um subconjunto de casos com vários vídeos em diferentes idades. Nós supomos que no autismo de início precoce, o manhês irá diminuir ao longo do tempo em termos de frequência relativa devido à falta de retorno interativo da criança. Este déficit na estimulação do desenvolvimento teria consequências, o que levaria, em particular, aos prejuízos sociais específicos e às alterações de linguagem encontradas em crianças autistas mais velhas (Zilbovicius et al., 2006). Na verdade, a prosódia do manhês está relacionada com as emoções da mãe e parece um autêntico marco da qualidade das interações entre pais e filhos. Na ausência de seus filhos, as mães não podem produzir "falso manhês" (Fernald e Simon, 1984): o leque de modificações prosódicas na fala da mãe é evocado apenas na presença da criança. Pelo contrário, as reações do bebê melhoram a amplitude dos contornos da prosódia da fala da mãe (Burnham et al., 2002). Mas quando um bebê parece indiferente, podem os seus pais continuarem a produzir o manhês? Se o manhês já não pode ser produzido por esses pais, a escassez de manhês parece ser um dos co-fatores no desenvolvimento de um círculo vicioso interativo. A Figura 3 mostra como o detector pode ser útil na realização de tal estudo. Em uma seqüência com duração de 12 segundos durante o qual ocorreu uma interação positiva (note o sorriso do bebê), o detector pode analisar automaticamente o sinal de som, características prosódicas e determinar se o manhês está sendo produzido.



**Figura 3:** Utilização do detector como uma ferramenta para o estudo do manhês em vídeos caseiros. Em uma sequência de 12 segundos de duração, durante os quais uma interação positiva ocorreu (ver o sorriso da criança), o detector de análise automática de sinal acústico, características prosódicas (Frequência,  $Hz$ , em azul; Intensidade,  $dB$ , em vermelho), falante, e de ocorrência do manhês. Paralelamente, as interações positivas existentes são informatizadas com a mesma escala temporal no banco de dados de vídeos caseiros de Pisa. A linha vertical vermelha indica o momento da foto do bebê (neste caso, 8 segundos após o início da seqüência).

#### 7.4. AS POSSÍVEIS APLICAÇÕES FORA DO CAMPO DO AUTISMO

Se as nossas hipóteses anteriores forem confirmadas (ou seja, se o decréscimo de manhês aparece como um co-fator no processo de patogenicidade), uma possível aplicação para crianças em risco de autismo pode favorecer o manhês parental. Como o manhês foi mostrado como efetivo para melhorar a receptividade e a aprendizagem em várias crianças deficientes (Santarcangelo e Dyer, 1988), seu uso poderia ser útil também para os enfermeiros, professores ou outros profissionais que trabalham no campo da saúde mental infantil fora do campo do autismo.

Como o manhês é um fator importante e uma ferramenta valiosa para estudar tanto as primeiras interações quanto o desenvolvimento da linguagem, podemos facilmente imaginar possíveis aplicações de sua detecção automática de vários projetos de investigação realizados ReVEL, v. 8, n. 15, 2010.

com um desenho experimental ou naturalista (por exemplo, Braarud e Stormark, 2008). Da mesma forma, a prosódia do manhês reflete o efeito emocional transmitido pela voz (Trainor et al., 2000). Como a análise manual tem consumido muito tempo, a detecção automática do manhês também pode ser útil para vários tipos de investigação sobre o discurso emocional (por exemplo, Beaucousin, 2007).

## 8. CONCLUSÃO

Usando recursos diferentes, segmentares e supra-segmentares, fomos capazes de discriminar automaticamente segmentos de manhês de segmentos de fala dirigida ao adulto, com baixos níveis de erros. Concluímos que o detector de manhês é uma ferramenta eficaz para estudar o manhês em vídeos caseiros, permitindo um estudo aprofundado de grandes amostras de seqüências.

## REFERÊNCIAS BIBLIOGRÁFICAS

1. ADRIEN, Jean-Louis, et al. Blind ratings of early symptoms of autism based upon family home movies. *Journal of the American Academy of Child and Adolescent Psychiatry*, n.32, 1993.
2. BARON-COHEN, Simon, et al. Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*, n. 168, 1996.
3. BEAUCOUSIN, Virginie; LACHERET, Anne; TURBELIN, Marie-Renée; MOREL, Michel; MAZOYER, Bernard; TZOURIO-MAZOYER, Nathalie. FMRI study of emotional speech comprehension. *Cerebral cortex*, n. 17, 2007.
4. BRAARUD, Hanne C.; STORMARK, Kjell Morten. Prosodic Modification and Vocal Adjustments in Mothers' Speech During Face-to-face Interaction with Their Two- to Four-month-old Infants: A Double Video Study. *Social Development*, Vol. 17, No. 4, 2008.
5. BRAZELTON, Terry B, et al. Early mother-infant reciprocity. In M. HOFER (Ed.). *Parent-infant interaction*. Amsterdam: Elsevier, 1975.
6. BURNHAM, Denis; KITAMURA, Christine; VOLLMER-CONNA, Uté. What's new pussycat: On talking to animals and babies. *Science*, n. 296, 2002.
7. CHAMAKA, Brigitte; BONNIAUA, Beatrice; JAUNAYB, Emmanuel; COHEN, David. What can we learn about autism from autistic persons? *Psychotherapy and Psychosomatics*, n. 77, 2008.

8. CLIFFORD, Sally; YOUNG, Robyn; WILLIAMSON, Paul. Assessing the early characteristics of autistic disorder using video analysis. *Journal of Autism and Developmental Disorders*, n. 37, 2007.
9. COHEN, David, et.al. Specific genetic disorders and autism: clinical contribution towards identification. *Journal of Autism and Developmental Disorder*, n. 35, 2005.
10. CONDON, William S.; SANDER, Louis W. Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science*, n. 183, 1974.
11. COOPER, Robin Panneton; ASLIN, RICHARD N. Preference for infant-directed speech in the first month after birth. *Child Development*, n. 61, 1990.
12. CORNUEJOLS, Antoine; MICLET, Laurent; KODRATOFF, Yves. *Apprentissage artificiel concepts et algorithmes*. Paris: Editions Eyrolles, 2002.
13. DUDA, Richard O.; HART, Peter E.; STORK, David G. *Pattern Classification*. Second edition, Wiley, 2000.
14. FERNALD, Anne. *Acoustic determinants of infant preference for "Motherese"*. In: Doctoral Dissertation, University of Oregon, 1982.
15. FERNALD, Anne. Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, n. 8, 1985.
16. FERNALD, Anne; KUHL, Patricia. Acoustic determinants of infant preference for Motherese speech. *Infant Behavior and Development*, n. 10, 1987.
17. FERNALD Anne, THOMAS, Simon. Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology*, 20, 1984.
18. GLENN, Sheila M.; CUNNINGHAM, Cliff C. What do babies listen to most? A developmental study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Developmental Psychology*, n. 19, 1983.
19. GOLDSTEIN, Michael H.; KING, Andrew P.; WEST, Meredith J. Social interaction shapes babbling: testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, n. 100, 2003.
20. GRIESER, DiAnne L; KUHL Patricia K. Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. *Developmental Psychology*, n. 24, 1988.
21. GROEN, Wouter B.; ORSOUW, Linda van; ZWIERS, Marcel; SWINKELS, Sophie; GAAG, Rutger Jan van der; BUITELAAR, Jan K. Gender in voice perception in autism. *Journal of autism and developmental disorders*, n. 38, 2008.

22. HIRSH-PASEK, Kathy, et al. Clauses are perceptual units for young infants. *Cognition*, 26, 1987.
23. KARZON, Roanne Gottlieb. Discrimination of polysyllabic sequences by one- to four-month-old infants. *Journal of Experimental Child Psychology*, n. 39, 1985.
24. KIM, Samuel; GEORGIU, Panayiotis; LEE, Sungbok; NARAYANAN, Shrikanth. Real-time emotion detection system using speech: Multi-modal fusion of different timescale features. *IEEE International Workshop on Multimedia Signal Processing*, 2007.
25. KEMLER NELSON, Deborah G., et al. How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, n. 16(1), 1989.
26. KLIN, Ami. Young autistic children's listening preferences in regard to speech: a possible characterization of the symptom of social withdrawal. *Journal of autism and Developmental Disorders*, n. 21, 1991.
27. KUHL, Patricia, et al. Cross-language analysis of phonetic units in language addressed to infants. *Science*, n. 277, 1997.
28. KUHL, Patricia. Colloquium Paper: A new view of language acquisition. *PNAS*, n. 97(22), 2000.
29. KUHL, Patricia. Human speech and birdsong: Communication and the social brain. *The Proceedings of the National Academy of Sciences*, n. 100 (17), 2003.
30. KUHL, Patricia. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5: 831-843. DOI:10.1038/nrn1533
31. KUHL, Patricia, et. al. Links between social and linguistic processing of speech in preschool children with autism: behavioral and electrophysiological measures. *Developmental Science*, n. 8(1), 2005.
32. KUHL, Patricia. Is speech learning 'gated' by the social brain? *Developmental Science*, n. 10:1, 2007.
33. KUNCHEVA, Li. *Combining pattern classifiers: Methods & algorithms*. Wiley, 2004.
34. LAZNIK, Marie-Christine; MAESTRO, Sandra; MURATORI, Filippo; PARLATO, Erika. Les interactions sonores entre les bébés devenus autistes et leurs parents. In: CASTAREDE Marie-France; KONOPCZYNSKI, Gabrielle (eds). *Au commencement était la voix*. Paris: Erès, 2005.
35. MAESTRO, Sandra, et. al. Study of the Onset of Autism through Home Movies. *Psychopathology*, n. 32, 1999.
36. MAESTRO, Sandra, et. al. Early Behavioral Development in Autistic Children: the first 2 Years of Life through Home Movies. *Psychopathology*, n. 34, 2001.

37. MAESTRO, Sandra, et. al. Attentional skills during the first 6 months of age in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, n. 41 (10), 2002.
38. MAESTRO, Sandra, et. al. Course of autism signs in the first year of life. *Psychopathology*, n. 38, 2005a.
39. MAESTRO, Sandra, et. al. How young children treat objects and people: an empirical study of the first year of life in autism. *Child psychiatry and Human Development*, n. 35 (4), 2005b.
40. MAESTRO, Sandra, et. al. A view to regressive autism through home movies. Is early development really normal? *Acta Psychiatrica Scandinavica*, n. 113, 2006.
41. MEHLER, Jacques; DUPOUX, Emmanuel. *Naître humain*. Paris: Odile Jacob, 1995.
42. MILGRAM, Maurice; ATLAN, Henri. Probabilistic automata as a model for epigenesis of cellular networks. *Journal of Theoretical Biology*, n. 21; 103(4), 2005.
43. MURATORI, Filippo; MAESTRO, Sandra. Autism as a downstream effect of primary difficulties in intersubjectivity interacting with abnormal development of brain connectivity. *International Journal for Dialogical Science*, n. 2 (1), 2007.
44. OSTERLING, Julie A.; DAWSON, Geraldine; MUNSON, Jeffrey A. Early recognition of 1-year old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, n.14, 2002.
45. PALOMO, Rubén; BELINCHÓN, Mercedes; OZONOFF, Sally. Autism and family home movies: a comprehensive review. *Developmental and Behavioral Pediatrics*, n. 27, 2006.
46. PEGG, Judith; FELDMAN, Janet; WERKER, Peter McLeod. Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, n. 15, 1992.
47. REYNOLDS, Douglas A. Speaker identification and verification using Gaussian mixture speaker models. *Speech Communication*, n. 17, 1995).
48. RINGEVAL, Fabien; CHETOUANI, Mohamed. Exploiting a Vowel Based Approach for Acted Emotion Recognition. *International Workshop on Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction*, 2008.
49. RUTTER Michael, LE COUTEUR Anne ; Catherine LORD. ADI-R: the autism diagnostic interview-revised. Los Angeles: Western Psychological Services. *Psychology and Psychiatry*, n. 48(12), 2003.
50. SAINT-GEORGES, Catherine; CASSEL, Raquel; COHEN, David; CHETOUANI, Mohamed; LAZNIK, Marie-Christine; MAESTRO, Sandra; MURATORI, Filippo. What ReVEL, v. 8, n. 15, 2010.

- the literature on family home movies can teach us about the infancy of autistic children: a review of literature. *Journal of Autism and Developmental Disorders*. (in revision).
51. SCHOPLER, Eric; REICHLER, Robert J.; RENNER, B. R. *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services. 1988.
  52. SANTARCANGELO, Suzanne; KATHLEEN, Dyer. Prosodic aspects of motherese: effects on gaze and responsiveness in developmentally disabled children. *Journal of Experimental Child Psychology*, n. 46, 1988.
  53. SCHULLER, Bjoern, et al. The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functionals. *Proceedings of InterSpeech*, 2007.
  54. SHAMI, Mohammad; VERHELST, Werner. An Evaluation of the Robustness of Existing Supervised Machine Learning Approaches to the Classification of Emotions (2007). An Evaluation of the Robustness of Existing Supervised Machine Learning Approaches to the Classification of Emotions. *Speech Communication*, n. 49 (3), 2007.
  55. SHORT, Andrew B.; SCHOPLER, Eric. Factors relating to age of onset in autism. *Journal of Autism and Developmental Disorders*, n. 18(2), 1988.
  56. STERN, Daniel, et. al. Vocalization in unison and alternation: Two modes of communication within the mother-infant dyad. *Annals of the New York Academy of Science*, n. 263, 1975.
  57. STONE, Wendy L.; HOFFMAN, Edward L.; LEWIS, Susan E.; OUSLEY, Opal Y. Early recognition of autism: Parental reports vs. clinical observation. *Archives of Paediatrics and Adolescent Medicine*, n. 148, 1994.
  58. SULLIVAN, Ann; KELSO, Jane; STEWART, Mark. Mothers' views on the ages of onset for four childhood disorders. *Child Psychiatry Human Development*, n. 20, 1990.
  59. SULLIVAN, Michelle; FINELLI, Julianna; MARVIN, Alison; GARRETT-MAYER, Elizabeth; BAUMAN, Margaret; LANDA, Rebecca. Response to joint attention in toddlers at risk for autism spectrum disorder: a prospective study. *Journal of Autism and Developmental Disorders*, n. 37, 2007.
  60. TRAINOR, Laurel J. ; AUSTIN, Caren M.; DESJARDINS, Renée N. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, n. 11, 2000.
  61. VOLKMAR, Fred R.; PAULS, David L. Autism. *Lancet*, n. 362, 2003.

62. WERKER, Janet F.; MCLEOD, Peter J. Infant preference for both male and female infant-directed talk: A developmental study of attentional affective responsiveness. *Canadian Journal of Psychology*, n. 43, 1989.
63. ZAKIAN, Alfred, et al. Signes précoces de l'autisme et films familiaux: une nouvelle étude par cotuteurs informés et non informés du diagnostic. *L'Encéphale*, n. 26, 2000.
64. ZILBOVICIUS, Monica, et. al. Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences*, n. 29(7), 2006.

**ABSTRACT:** Autism is a well-defined clinical syndrome after the second year of life, but information on autism in the first two years of life is still lacking. The study of home videos has described children with autism during the first year of life as not displaying the rigid pattern typical of later symptoms. Therefore, developmental/environmental factors are claimed in addition to genetic/biological ones to explain the onset of autism during maturation. Here we describe (1) a developmental hypothesis focusing on the possible implication of motherese impoverishment during the course of parent/infant interactions as a possible co-factor; (2) the methodological approach we used to develop a computerised algorithm to detect motherese in home videos; (3) the best configuration performance of the detector in extracting motherese from home video sequences (accuracy=82% on speaker-independent versus 87.5% on speaker-dependent) that we should used to test this hypothesis.

**KEYWORDS:** Motherese; Autism; Computerised detector.

Recebido no dia 05 de junho de 2010.

Artigo aceito para publicação no dia 30 de julho de 2010.