

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e Computação
Departamento de Semicondutores, Instrumentos e Fotônica

Interface Gestual Ultrassônica

AUTOR : GILBERTO MENDES

ORIENTADOR : PROF. DR. FURIO DAMIANI

Comissão Julgadora:

Furio Damiani – DSIF / FEEC / UNICAMP
Jônatas Manzolli – NICS / UNICAMP
Peter J. Tatsch – DSIF / FEEC / UNICAMP

Tese apresentada à Faculdade de Engenharia Elétrica e Computação da Universidade Estadual de Campinas (FEEC-UNICAMP) como parte dos requisitos exigidos para obtenção do título de Mestre em Engenharia Elétrica

Campinas – São Paulo – Brasil

Dezembro de 1999

Dedico este trabalho à minha família e aos bons amigos que, durante a realização deste trabalho, deram apoio, paciência e crítica nas doses certas.

Agradecimentos

Agradeço ao Prof. Dr. Furio Damiani pela orientação e apoio incondicionais, acrescidos a uma boa dose de estímulo às novas idéias ; ao Prof. Dr. Jônatas Manzolli (NICS-UNICAMP), por ajudar a tornar útil este trabalho em pesquisas futuras; ao Prof. Dr. Julio Cezar Adamowski (Dep^{to}. de Engenharia Mecânica – Esc. Politécnica/USP), pelo incentivo e inspiração ; aos amigos Farley Simon Mendes Nobre e Pedro Monerat, pela receptividade e pela participação constante no amadurecimento do trabalho ; ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro; e aos meus familiares e amigos que não permitiram que minha ausência no período de condução deste trabalho atrapalhasse os laços construídos anteriormente.

Sumário

Dedicatória	ii
Agradecimentos	iii
Sumário	iv
Resumo	v
Abstract	vi

Capítulo 1 - Introdução

1.1 Interface Homem-Máquina	1
1.2 Objetivos deste Trabalho	6
1.3 Organização da Tese	7

Capítulo 2 - Medição de Posição

2.1 Métodos de Medição de Posição	8
2.2 Sistemas de Sonar e Exemplos na Natureza	10
2.3 Método de Medição de Posição Proposto	17
2.4 Sistema de Transmissor-Sensores	29
2.5 Aquisição de Dados e Cálculo de Posição em Três Dimensões	36

Capítulo 3 - Identificação de Gestos

3.1 Delimitação de Gestos	44
3.2 Redes Neurais e Classificação de Sequências	44
3.3 Pré-processamento e apresentação de dados à Rede Neural	53
3.4 Métodos de Treinamento, Topologias Usadas e Resultados	59

Capítulo 4 - Conclusões e Perspectivas Futuras

4.1 Sobre a Medida de Posição por Ultra-Som	66
4.2 Sobre o Reconhecimento de Gestos	68

Apêndice A – Listagem do Software Desenvolvido	70
---	-----------

Referências Bibliográficas	80
-----------------------------------	-----------

Resumo

Com o desenvolvimento de sistemas multimídia, cresce a necessidade de formas de interação mais naturais entre Homem e máquina. Este artigo descreve um dispositivo de entrada de comandos a partir de gestos humanos, reconhecidos por uma rede neural treinada pelo usuário para se adequar a um ambiente de trabalho específico.

A interface proposta utiliza um pequeno transmissor de ultra-som na mão do usuário, apontado para um conjunto de sensores próximo ao computador em uso. A trajetória de um gesto feito pelo usuário é amostrada ponto a ponto, usando uma série de temporizadores digitais integrados numa EPLD, pré-processado e apresentado à rede neural para interpretação.

Abstract

The development of multimedia systems ensues the creation of man-machine interfaces that allow a more natural interaction. This paper describes a command input device able to recognize user gestures. Recognaissance is done by a user trained neural network, to fit specific working media needs.

The device uses a handheld ultrasonic transmitter, steered to a set of sensors close to the computer in use. The device movement is time sampled using a series of digital timers integrated in one EPLD, preprocessed and fed to the neural network for interpretation.

Capítulo 1 - Introdução

1.1 Interface Homem-Máquina

A interação do Homem com os computadores (ou, de forma mais genérica, com as máquinas), está longe de ser comparada com a forma corriqueira de comunicação entre os Homens e nem é derivada da mesma, obrigando a adaptação do Homem aos padrões de operação da máquina [40] [46], se quiser se fazer entendido.

A comunicação entre seres humanos é normalmente composta de sons (com variações de entonação e de dinâmica), expressões faciais, posições (estáticas) do corpo, movimentos do corpo ou de partes deste, o toque (abraços, socos, apertos de mãos, etc.), e o cheiro. Estas formas de expressão podem ser usadas simultaneamente para se complementar, para tornar mais clara uma mensagem. Além disso, algumas das formas de expressão são específicas para representação artística (teatro) ou produzir música [46].

Neste trabalho, é proposta uma forma de interface entre o Homem e uma máquina, seja um computador pessoal, um controlador industrial, ou um instrumento musical eletrônico. O trabalho aqui descrito vem a expandir o conjunto de dispositivos já existentes, como os teclados, os apontadores eletrônicos (mouse, tablete gráfico), e pode complementar a função de transmitir comandos à máquina através de gestos no espaço. Foram desenvolvidos um sistema eletrônico (*hardware*) e programas de computador (*software*) que executam a aquisição de gestos em três dimensões. Sua aplicabilidade foi testada na prática pelo reconhecimento de gestos escolhidos por um usuário como comandos específicos, programados por meio de treinamento de redes neurais artificiais.

1.1.1 Gestos

A maioria dos seres humanos se utiliza de gestos [40] para:

- completar o sentido de uma mensagem falada, ou
- conter por si só a mensagem completa, ou
- quando alguém não consegue ou não quer falar, ou
- quando se quer comunicar com alguém sem que outro próximo perceba a mensagem.

A definição de gesto pode ser restrita aos movimentos de partes do corpo com algum sentido ligado à cultura de quem os realiza. Pode-se definir o gesto de forma mais ampla, incluindo posições (estáticas em um intervalo de tempo) de todo o corpo ou de suas partes. A forma de aquisição de dados de um gesto implementada neste trabalho condiz com a definição

mais ampla, mas é possível desconsiderar as posições estáticas em outras aplicações (outros *softwares*) usando o mesmo sistema eletrônico.

Os gestos são tão vinculados à comunicação humana que quase sempre nem percebemos quando os fazemos. Este fato os coloca, sem dúvidas, como parte de uma cultura ligada à forma de se comunicar. Não são essenciais para se fazer entender (as pessoas conseguem se comunicar bem por telefone), mas são complementos riquíssimos para a comunicação quando estão presentes [40].

Na grande maioria dos casos, porém, quando alguém precisa transmitir um comando para uma máquina, deixa de lado toda a riqueza de sua cultura e se adapta ao protocolo de entrada de comandos específicos da máquina [40]. Ainda não podemos chegar perto de um telefone e dizer a ele o que queremos que ele faça, muito menos gesticular para pedir urgência. Já temos exemplos de máquinas que interpretam de comandos vocais, e estão sendo testadas interfaces homem-máquina que complementam comandos tradicionais de teclado ou mouse pela adição de comandos vocais concomitantes. Mesmo assim, não se alterou o fato de que o teclado e o mouse, sendo os dispositivos mais difundidos em uso para comandar um computador, fazem com que a posição de trabalho e a dinâmica de movimentos do usuário sejam determinadas pela sua forma de operação. Estas restrições são dignas de reavaliação, pois chegam a causar lesões nos usuários que passam muito tempo se sujeitando a elas.

Uma estratégia conveniente de delimitação de gestos ao capturá-los por meios eletrônicos deverá oferecer o máximo de flexibilidade de uso, para se aproximar da forma com que o Homem se comunica normalmente. Pode se optar por captar gestos com começo, meio e fim, como movimentos delimitados por alguma forma de segmentação dinâmica reconhecível por um observador (pausas entre gestos, por exemplo). Outra opção interessante seria ficar continuamente descrevendo os últimos movimentos por funções matemáticas para a extração de características, como a rapidez de movimento, a cadência do mesmo, as curvaturas dos contornos. Isto permite a interpretação e a reação conveniente a uma expressão gestual mais livre, menos atrelada a padrões. Uma conjunção destas duas formas de interpretação de gestos é usada em vários trabalhos [63,20], extraindo características de etapas de um evento, juntando as partes e tentando o reconhecimento do conjunto como um comando gestual.

As mãos de quem conversa com entusiasmo estão em contínuo movimento, mas quando o objetivo é comandar uma ação, os gestos são normalmente restritos a alguns padrões para possibilitar seu entendimento (uma pessoa ensinando um caminho representa com as mãos as direções que se deve tomar, cada uma com um gesto estático). Este fato diminui bastante a complexidade de uma estratégia de interpretação de gestos por computador, quando estes significarem comandos padronizados. Além disso, entre um gesto e outro há sempre movimentos de transição que se deve desprezar. O entendimento da porção gestual do que um usuário de computador expressa, depende da separação das porções úteis do movimento das mãos, para traduzir o que se deseja que o computador faça.

Uma forma de separar os gestos é ficar atento às pausas, delimitando um gesto entre duas pausas pronunciadas. É uma regra baseada na observação das pessoas, que pode ser

aplicada em sistema de reconhecimento de gestos em tempo real. As pausas seriam mais um gesto a ser reconhecido, sendo usadas para delimitar seqüências sem um número fixo de pontos, que podem ser re-amostradas com um número de pontos padronizado, separados como um gesto completo e entregues a uma rede neural. Esta, por sua vez, tentaria reconhecê-lo como um dos gestos válidos e associá-lo a um comando pré-programado.

Uma outra forma de usar a interface aqui proposta, alternaria seu uso como apontador (com a vantagem da terceira dimensão) e como entrada de gestos. A informação de um trio de coordenadas de um ponto poderia ser usada de imediato para posicionamento de um cursor na tela de um computador servindo para operar uma interface gráfica, apontando, selecionando, arrastando ícones e outros objetos, da mesma forma que um mouse. Pode operar assim indefinidamente até que se faça um gesto determinado (delimitado por pausas, por exemplo, e simples de ser reconhecido) que muda a forma de interpretar os movimentos. Após a execução de cada gesto completo subsequente poderia se tentar seu reconhecimento em tempo real por uma rede neural (ou usando outros recursos para reconhecimento de padrões) e como consequência partir para a execução de uma tarefa associada a este gesto (1 gesto = 1 comando). Os gestos seriam interpretados como comandos específicos para uma dada aplicação, se fossem reconhecidos. Um destes gestos comandaria a volta à função de apontador. É desta forma que se imaginou inicialmente a utilidade prática deste projeto, com a ajuda de *softwares* interativos, para explorar as possibilidades da interação com a máquina por meio de gestos.

1.1.2 Divulgação Científica

A necessidade de criação de formas alternativas de interação entre uma máquina e seu usuário que evitem os inconvenientes das interfaces que hoje têm seu uso mais difundido motivam a investigação de dispositivos eletrônicos dedicados à determinação das posições (estáticas) do corpo ou de partes específicas deste, ou da movimentação dos mesmos.

Vários trabalhos têm sido publicados como tentativas de satisfazer esta necessidade, em sua maioria usando câmeras de vídeo e algoritmos sofisticados de processamento de imagens, de alto custo computacional [6,32,40,45,63]. Alguns trabalhos foram realizados utilizando ondas acústicas para localizar pontos do espaço [1,7,11,14,35,39,49,53], assim como neste trabalho. Alguns têm feito a conjugação de vários princípios de medição de posição [6,35], com as vantagens advindas da fusão de sensores, mas com barreiras de realização em equipamentos acessíveis ao público de forma ampla, como seus custos (muito altos) e o incômodo gerado pelo seu uso (peso, tamanho).

A classificação dos diversos tipos de dispositivos que fazem a leitura de movimentos humanos pode ser feita a partir da escolha do posicionamento relativo entre transmissores e sensores (quais ficam no corpo do usuário, quais ficam externos), pelo meio usado para sensoriamento (acústico, eletromagnético, ótico, mecânico, etc.), pelo tipo de interação psicomotor envolvida (músculo-esquelético, neuro-motor, simbólica) [45] ou por outras características ligadas a estudos mais específicas de alguma área de estudos, como por

exemplo a classificação de instrumentos musicais virtuais pela forma de expressão, ou de controle dos sons emitidos (gestos, dança, movimentação de um objeto) [46].

1.1.3 Dispositivos Apontadores

Os sistemas de Interface Gráfica com o Usuário (GUI - Graphical User Interface), tão difundidos nos sistemas de computação pessoal e profissional, devem grande parte de sua praticidade aos dispositivos apontadores. A operação destas interfaces pode ser vista como uma sequenciamento de ações do usuário que aponta e movimenta objetos em uma tela gráfica. O controle é feito de forma indireta, usando dispositivos de entrada (o mais difundido, sem dúvida, é o mouse) que traduzem movimentos em códigos de computador. Estes códigos determinam, então, o ponto de atuação (em duas dimensões) do usuário na interface gráfica em uso, que oferece um retorno visual para constante reajuste deste ponto [32].

Uma das características deste tipo de controle normalmente encontrada nos aplicativos é o posicionamento fixo, muitas vezes preciso, dos objetos no espaço de trabalho, o que não corresponde às características da maioria dos movimentos de interação do Homem com seu semelhante e com o ambiente. Em muitas situações, usamos um dedo para apontar, sem muita precisão, na direção de um objeto ao qual nos referimos, contando com a capacidade de discernimento de quem está ouvindo ou completando a informação com o que falamos, por exemplo: “ali, bem no canto direito, acima daquele ponto azul” . Os gestos exprimem ou complementam idéias [40] (acenar para se despedir, um V para dizer Paz e Amor, e outros exemplos) e não contêm, obrigatoriamente, noções de precisão. Não seria a mesma coisa se tivéssemos que apontar para os objetos sempre com precisão de milímetros. Portanto, são raros os movimentos que o usuário faz para operar seu computador que correspondem à sua bagagem cultural [40]. O uso de gestos para comunicação pode herdar esta limitação imposta pelos aplicativos atualmente em uso, adequados para o uso com os dispositivos apontadores existentes, ou pode provocar a adaptação dos aplicativos a uma forma menos restritiva de interação com a máquina.

A interpretação de um gesto do usuário pode ser usada para provocar reações diferenciadas, dependendo do contexto, compondo um conjunto de eventos que controlam um sistema interativo. Podemos encontrar em publicações científicas alguns trabalhos de pesquisa de interfaces com o usuário que associam várias formas de expressão e que dimensionam as interações entre as mesmas, adequando-se a cada aplicação. O exemplo visto em [32] envolve várias tecnologias, incluindo formas alternativas de interação com a máquina (entre estas, os gestos) de forma a torná-los colaborativos; este sistema também está preparado para gerenciar bases de conhecimento por inteligência artificial para caracterizar de forma adaptativa a dinâmica de integração entre as formas de expressão do usuário.

Com o acréscimo da capacidade de entender gestos, um computador tem seu espectro de aplicações expandido [40]. Pode, por exemplo, auxiliar o trabalho de desabilitados, sendo enriquecida a interação usuário-computador. Pode tornar possível a execução de instrumentos

musicais virtuais, ou permitir a interferência de forma dinâmica na composição algorítmica de música com expressões gestuais (veja detalhes no último capítulo).



Figura 1.1

O trabalho aqui proposto oferece ao usuário de uma máquina uma forma de apontar para objetos em uma interface gráfica, alternativa aos apontadores conhecidos, acrescentando maior liberdade de movimentos e uma dimensão a mais (3ª dimensão). Foi testada a sua utilização com uma das aplicações possíveis, que permite ter gestos corriqueiros entendidos pelo computador, na forma de comandos completos adequados a um contexto. Tais gestos podem também complementar um comando por voz (fig. 1.1). Imagine-se dizendo ao computador: “ponha este objeto ali” (fig. 1.2), com a ajuda de gestos [40].

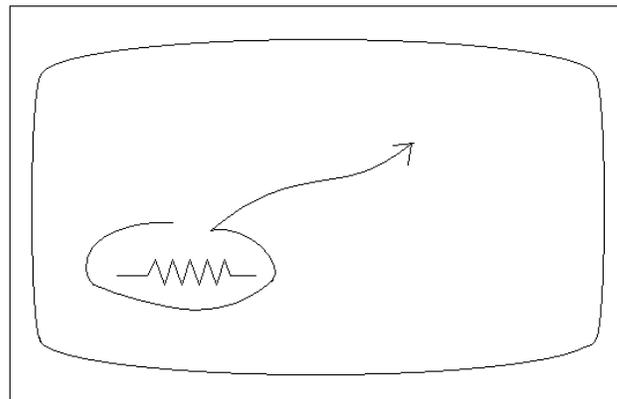


Figura 1.2 - “ponha este objeto ali ” ou “put that there”/MIT

1.2 Objetivos deste Trabalho

Este trabalho tem como objetivo desenvolver um sistema eletrônico para medir a localização de um pequeno transmissor em três dimensões e testar (com desenvolvimento de programas para computadores pessoais) a aquisição e a interpretação de seqüências destas medidas como integrantes de um gesto. O sistema eletrônico se compõe de um conjunto de transmissor-sensores de ondas sonoras na faixa de ultra-som, circuitos para condicionamento destes sinais e para medições de tempo de propagação de onda, e interface para leitura por um computador. Os programas devem executar a aquisição e a organização de dados em arquivos, cálculos de distâncias, treinamento e avaliação de redes neurais artificiais para o reconhecimento de gestos como comandos.

O transmissor deve ser compacto o suficiente para ser fixado no dedo indicador de um usuário (por exemplo) e não deve haver a necessidade de contato com a máquina por meio de fios. O alcance deve ser de alguns metros, para permitir a execução de gestos de boa amplitude, a uma distância razoável do computador.

O conjunto de sensores deve ser em número suficiente para medições em três dimensões e dispostos de forma a facilitar a tarefa de calcular as distâncias até o transmissor, podendo ser coplanares ou não.

O sistema deve ser concebido de forma a ser viável sua utilização em sistemas de baixo custo, portanto deve usar componentes baratos e de fácil aquisição. Este critério tem como objetivo tornar este trabalho reproduzível para instrumentalizar outros centros de pesquisa que necessitem de investigação detalhada de movimentos em tempo real.

A medição de localização deve ser feita a uma taxa (medições por segundo) que possibilite seu uso para captura de gestos, vistos como seqüências de pontos de um percurso realizado pelo transmissor anexado a um dedo do usuário.

Os programas para aquisição e organização de dados em arquivos serão desenvolvidos em linguagem C, devido à sua portabilidade, para se adaptar a outros sistemas computadorizados. Os programas para treinamento e avaliação de redes neurais serão desenvolvidos em Matlab, por oferecer bons recursos para interpretação e visualização de resultados de processamento matemático.

Este trabalho foi concebido para fazer parte de outras atividades de pesquisa. Servirá para coletar uma base de conceitos sobre o assunto, prevendo o desenvolvimento de estratégias mais complexas de aquisição e interpretação de gestos em aplicações futuras. É sugerida a pesquisa de outras formas de uso com restrições menos rígidas, mais confortáveis para o usuário, que ampliem a funcionalidade de máquinas.

1.3 Organização da Tese

Este texto se inicia com uma apresentação das motivações para a escolha do tema e dos rumos tomados ao longo do trabalho de pesquisa. Em seguida, no capítulo 2, foi feito um apanhado sobre as formas de medição de posição estudadas para que se fizesse uma escolha criteriosa, e são detalhados o método utilizado e a interface gestual desenvolvida. Uma das aplicações possíveis desta interface, a identificação de gestos, foi explorada para testar sua praticidade e eficiência. Uma rápida explanação sobre as redes neurais utilizada no treinamento e no reconhecimento de gestos foi incluída para entendimento de sua atuação no contexto do trabalho. Os resultados parciais e finais obtidos no reconhecimento de gestos são então apresentados com uma análise das limitações e das melhorias possíveis para adequação da interface às outras aplicações. O texto é finalizado com as conclusões a respeito da interface gestual ultrassônica e da aplicação realizada, sugerindo outras aplicações e as modificações possíveis para expansão de sua aplicabilidade.

Capítulo 2 - Medição de Posição

2.1 Métodos de Medição de Posição

2.1.1 Localização de objetos por ondas acústicas

Os fatores que diferenciam os sinais encontrados nos diversos sistemas de localização estão ligados ao ambiente em investigação e às características do objeto a localizar. Estes fatores são determinantes na escolha dos subsistemas eletrônicos de transdução, condicionamento, processamento dos sinais e apresentação dos resultados. Para medições no vácuo, por exemplo, não se poderia empregar ondas acústicas, que se propagam apenas em meios materiais. Um submarino, para evitar de ter sua presença percebida, não pode emitir nenhum sinal acústico para localizar o inimigo, portanto o faz pela recepção e processamento dos sinais acústicos emitidos pelos motores dos mesmos e pelo deslocamento de água.

São várias as formas de se determinar a direção e/ou a distância de objetos a partir de ondas acústicas. Quanto à geração das ondas acústicas, podem-se classificar os sistemas entre os de onda contínua e os de onda pulsada. Por outro lado, pela forma de recepção das ondas, temos os métodos de transmissão-recepção, associados muitas vezes a alguma forma de modulação, e os métodos de pulso-eco.

Nos métodos de onda contínua, a distância entre o transmissor e o sensor é medida pelas mudanças de fase entre a onda transmitida e a recebida. Por outro lado, nos métodos de onda pulsada, a distância é avaliada a partir da medida do tempo que um trem de pulsos de curta duração transmitido demora para se deslocar entre a transmissão do sinal e a recepção do mesmo.

Para se medir a posição, em duas dimensões, a partir de sensores em um plano, pode-se usar uma conjugação do método de onda contínua com o modo de transmissão-recepção, como sugerido em [49]. Naquela implementação, a posição de um transmissor localizado no objeto é computada a partir da diferença de fase entre ondas contínuas recebidas por dois pares de sensores (um par por eixo, X e Y) diretamente, e não entre transmissor e receptor. Um método similar associa um receptor ao objeto e usa técnicas de modulação de frequências [49].

No presente trabalho, é proposto um sistema de múltiplos sensores coplanares, para localização, em três dimensões, de objetos no espaço anexados a um transmissor de sinais ultra-sônicos e infra-vermelhos, de baixo custo. É usada uma conjugação de transmissão de ondas pulsadas com o modo de transmissão-recepção. A detecção de sinais na recepção é implementada com detecção de limiar de envoltória. A propagação de ondas diretamente de transmissor para receptor proporciona baixíssimos níveis de ruído e distorções comparados

aos que se encontram nos sistemas de pulso-eco e nos sonares para ambiente marinho, que foram a inspiração inicial para este trabalho.

Para se medir a posição usando vários sensores em um plano, a abordagem mais usada conjuga o método de ondas pulsadas e o modo pulso-eco, através de triangulação, a partir das distâncias absolutas entre um transmissor e mais de um receptor. Uma das dificuldades deste método é o baixo nível dos sinais de eco, que obriga ao uso de amplificadores de precisão e alto ganho, tornando crítica a relação sinal/ruído. Para a avaliação precisa, em tempo real, do tempo de propagação, é necessário o uso de algoritmos de processamento digital, de custo computacional significativo, para discriminar sinais na presença de ruído. Muitos trabalhos têm sido publicados com bons resultados neste campo de pesquisas, como visto em [7,56,52]. Podem-se implementar processadores com circuitos integrados dedicados, para alcançar um bom desempenho, mas condenando o critério de minimização de custos.

Outro exemplo de processamento de sinais de múltiplos sensores vem da adaptação do sistema sensorial dos morcegos para a caça noturna. Eles usam um sofisticado sistema de localização por ondas acústicas, do tipo pulso-eco, com frequência variável e diversos modos de interpretação dos sinais, além da conjugação de duas ou mais destas informações (correlatores). O método mais amplamente utilizado para medir distâncias por instrumentos eletrônicos deriva deste sistema dos morcegos, mas com apenas um modo de interpretação dos sinais. Um pulso (ou uma pequena quantidade de pulsos) é transmitido em direção a um anteparo, a onda reflete no mesmo e propaga de volta em direção ao instrumento. A onda refletida é então detectada por um receptor, determinando o instante de chegada da mesma. Para obter a distância (em metros) até o anteparo, o tempo total de propagação da onda é medido (em segundos), dividido por dois (pois é medido o tempo de ida e volta) e multiplicado pela velocidade do som no ar (em metros/segundo). Com as simplificações impostas por limitações de custo, surgem limitações para seu uso, por exemplo:

- em linha de visada bem diretiva, por que a concentração do ângulo de emissão da onda reduz a potência requerida do transmissor e a concentração do ângulo de recepção melhora a sensibilidade do receptor;
- com anteparos planos, se quiser atingir o alcance máximo, pois as superfícies curvas ou muito rugosas provocam a dispersão das ondas;
- com materiais mais densos, como metal, madeira, concreto, vidro, borracha, nos quais a reflexão chega próximo de 100 % , ao passo que roupas, cortinas, luvas de lã, absorvem as ondas acústicas, dificultando a medida de distância;
- com alvos imóveis, em alguns casos, devido ao tempo de processamento necessário, por exemplo, para extração de médias com o objetivo de filtrar as variações entre medidas (devido à baixa resolução imposta pelo comprimento de onda, ou à adição de ruídos pelo próprio circuito eletrônico) ;
- outras limitações mais específicas, como o tamanho mínimo do objeto, distância mínima, nível de ruído no ambiente, temperatura de operação, etc..

2.1.2 Localização de objetos por outros métodos

Usando princípios de funcionamento bem semelhantes aos sistemas de localização por sinais acústicos, há sistemas que usam ondas eletromagnéticas, como os radares convencionais para localização de alvos para fins militares, que usam a faixa de microondas, os “radares” de luz infra-vermelha, as câmeras de imagens térmicas para faixas de comprimentos de onda de milímetros, outras para a faixa de microondas, radares a laser. As diferenças entre eles começam nas frequências de operação, e não vão muito longe, mudando a interpretação dos sinais, mais precisos e quantitativos em uns (de menor frequência), mais abrangentes e qualitativos em outros (de maior frequência).

Existem sistemas de múltiplos sensores que combinam grandezas diferentes, usando as informações de uma para complementar as da outra. Por exemplo, em um avião de combate, uma câmera para imagens térmicas é usada para localização de um alvo com pontos quentes (turbinas de outros aviões, outras máquinas), indicando a sua direção, posteriormente um radar a laser obtém a informação de distância e do tamanho do alvo. Neste caso, o radar a laser dificilmente conseguiria localizar um alvo no raio de 360°, pois tem um raio muito estreito. Por outro lado, ele complementa a informação de imagem térmica que não consegue avaliar distâncias. O radar a laser também ajuda a confirmar a presença de um alvo apontada pela imagem térmica, evitando-se assim alarmes falsos. Esta é outra vantagem da conjugação de sensores, a confirmação de medidas, mesmo com faixas de atuação diferentes [8] .

A conjugação de imagens de vídeo capturadas por duas câmeras de vídeo permitem extrair de uma cena as informações de profundidade da mesma forma que na visão humana, e são chamadas de visão estéreo. A informação de distância é extraída de procedimentos de focalização, e é proporcional ao ângulo entre as câmeras quando centralizam o mesmo objeto.

2.2 Sistemas de Sonar e Exemplos na Natureza

2.2.1 Definição e Histórico

Os sistemas de SONAR e RADAR são bastante conhecidos, mesmo fora da comunidade científica (por aparecerem freqüentemente em filmes) como instrumentos para localização de objetos sem que haja contato com o mesmo. O termo SONAR é um acrônimo para "*Sound Navigation and Ranging*" , e é definido originalmente como "método ou equipamento para determinação, por meio de ondas sonoras, através da água, da localização e/ou da natureza de objetos no mar" [36]. Há registros do uso de sonares desde o início do século XX. Dadas as semelhanças encontradas com o sonar a que se refere definição acima, o sistema de orientação usado pelos morcegos e golfinhos, e as múltiplas aplicações de instrumentos eletrônicos para localização de objetos por propagação de som no ar, também tem sido referenciados como sonares, sem a restrição quanto ao meio de propagação presente na definição.

Os eventos históricos que mais marcaram a evolução do uso de ondas sonoras para localização de objetos foram [64] :

- # 1635 – Primeiros registros de medida da velocidade do som no ar.
- # 1822 – Daniel Colladen, um físico suíço, usou um sino para determinar a velocidade do som na água
- # 1880 – Os irmãos Curie descobriram o efeito piezoelétrico de certos cristais, que seriam usados mais tarde nos transdutores acústicos
- # 1912 – O naufrágio do navio Titanic por falta de um alerta mais adiantado da presença de um iceberg na sua rota, somado à necessidade de auxílio à navegação e localização de embarcações inimigas por submarinos na 1ª Guerra Mundial, motivou o início do desenvolvimento de sistemas de sonar.
- # 1928 – Sergei Y. Sokolov desenvolve detector de falhas em peças de metal para testes não destrutivos.
- # 1942 – Karl T. Dussik localiza tumores no cérebro pela propagação de ultra-som através do crânio, com um transdutor de cada lado.
- # 1949 – John J. Wild usa sinais de eco de ultra-som para diagnósticos de doenças no intestino e nos seios, inicialmente sem a produção de imagens, comparando os sinais de eco em vários pacientes, usando um sistema de radar adaptado para ondas sonoras, medindo centímetros no lugar de quilômetros. Três anos depois ele produz imagens de tumores em um osciloscópio adaptado

Nos anos '50 houve um grande avanço nos circuitos eletrônicos desenvolvidos para tornar os equipamentos de diagnóstico médico, de Sonar e de testes não destrutivos viáveis comercialmente, resultando em geradores de pulsos mais estreitos em frequências maiores, amplificadores mais sofisticados, melhorando a sensibilidade e a resolução das medidas. Desde então os avanços das tecnologias de circuitos eletrônicos e dos computadores melhoram de forma acelerada a qualidade das medidas e a comodidade de uso dos equipamentos nas três áreas mencionadas que usam ultra-som.

2.2.2 Classificação e utilização

Os sistemas de sonar são classificados, quanto à origem dos sinais captados, em ativos ou passivos. Os ativos emitem um sinal acústico de curta duração, de faixa estreita de frequências, que é refletido por objetos e tem seu retorno captado de forma seletiva (apenas na frequência emitida e em direções de interesse programáveis). Os passivos recebem sons em uma faixa larga de frequências produzidos pelos objetos de interesse, como ruído de motores, perturbações no fluxo de água, transmissões de sonares ativos de embarcações militares [36].

A escolha entre o uso de sonar passivo ou ativo, em aplicações militares, se dá por motivos táticos. Um submarino se baseia em um sonar passivo para detectar a presença, a direção e a classificação de um possível alvo, de forma que, não emitindo nenhum sinal acústico, se coloca em posição de ataque (ou seja, em distância menor que o alcance de suas armas) sem ser percebido. Uma vez em posição de ataque, o submarino emite um único pulso

de onda sonora para obter uma informação mais fina sobre a distância até o alvo. Uma embarcação de superfície provoca muito ruído na água ao se deslocar, tornando favorável sua detecção por sonares passivos. Ao mesmo tempo, tais embarcações ruidosas e visíveis acima da superfície da água não têm por que usar sonares passivos, optando-se então pelo uso dos ativos, que oferecem algumas vantagens, como a limitação da faixa de frequências de interesse, possibilitando o uso de filtros mais seletivos na recepção e eliminando a necessidade de conhecimento das emanações acústicas dos tipos de alvos de interesse. [9].

Nas décadas de '80 e '90, as investigações conduzidas pelos órgãos militares dos países do primeiro mundo se concentram em desenvolver a síntese de sinais muito mais elaborados para transmissão pelos sonares ativos, com o intuito de melhorar a capacidade de detecção de alvos e possibilitar sua classificação. Este desenvolvimento tem sido viabilizado pela disponibilidade de processadores digitais de sinais cada vez mais rápidos.

Além das utilizações para fins militares, os sonares são usados comercialmente para localização de cardumes, mapeamento de relevo do fundo do mar, auxílio à navegação, comunicação através da água, além de estudos sísmicos. Sistemas semelhantes são usados para a localização de lençóis de petróleo, empregando ondas acústicas com frequências muito inferiores.

Os morcegos usam um sonar ativo para localizar e capturar insetos no ar, além de se orientar nos vôos noturnos. Eles usam de processamento intenso na análise dos sinais de eco e na reação rápida às informações extraídas desta análise. Os golfinhos usam um sistema semelhante de sonar ativo, enquanto mariposas usam sonares passivos [41].

Os sistemas de sonar, em sua grande maioria, empregam conjuntos de sensores em forma de matriz e/ou a captação de sinais por sensores em movimento, com o objetivo de aumentar a diretividade do seu campo de sensibilidade, extraíndo apenas os sinais vindos de uma direção específica e atenuando os sinais de outras direções provocados por reflexões indesejadas. Esta técnica, conhecida como *beamforming* [11,15,27,30,36,38,50,54], com origem no sistema de orientação de antenas de rádio, permite a mudança da direção de maior sensibilidade por controle digital, e será mais detalhada mais adiante.

2.2.3 Propagação de ondas acústicas

As ondas acústicas são vibrações mecânicas que se propagam em um meio material, devido ao deslocamento de cada partícula do meio oscilando em torno de sua posição de equilíbrio, sem que haja um movimento residual de conjuntos destas partículas em nenhuma direção. Os movimentos das partículas causam uma série de regiões de compressão e descompressão, causando mudanças correspondentes na densidade e no índice de refração de luz locais no meio de propagação. Se a fonte de ondas acústicas for pontual e omnidirecional (pressão acústica igual em todas as direções), será originada uma série de ondas esféricas, representadas na fig. 2.1 (em duas dimensões, como um corte passando pelo centro das ondas esféricas e da fonte de ondas). Pode-se observar a representação das diferenças de pressão ao longo da onda pela graduação dos tons de cinza.

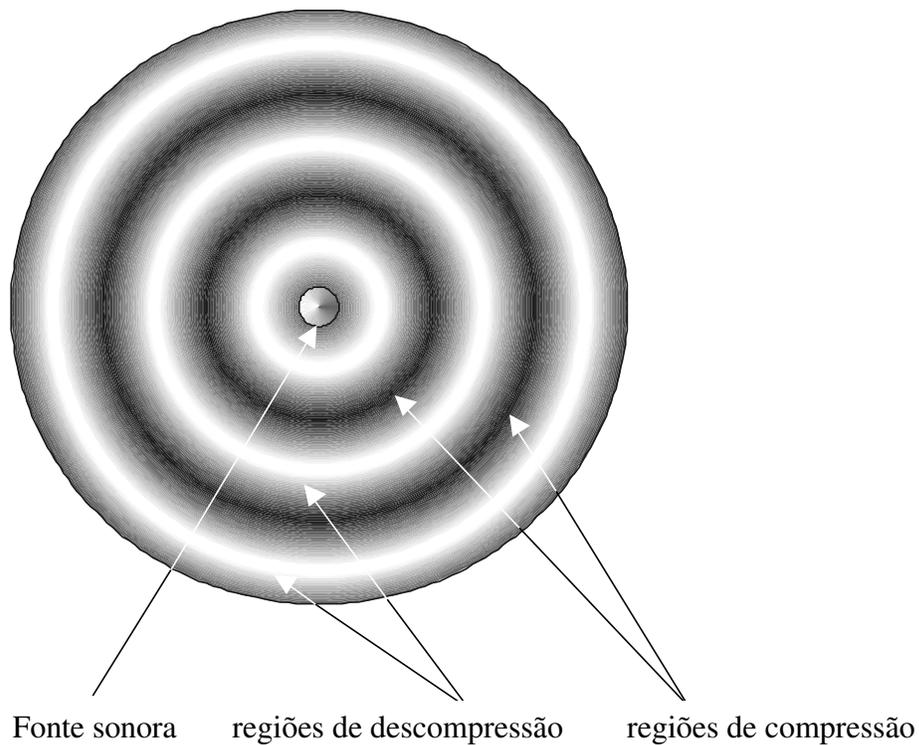


Figura 2.1 – Propagação de ondas acústicas

A velocidade das ondas acústicas em um meio varia com a temperatura, e é igual ao produto entre a frequência e o comprimento de onda (veja mais adiante, neste mesmo capítulo, equações que estabelecem as relações entre estas grandezas e as simplificações feitas para uso neste trabalho). Para a propagação do ultra-som no ar, podem ser feitas algumas aproximações sobre a velocidade de propagação:

- é independente da pressão atmosférica
- é maior em uma atmosfera úmida, pois os pesos moleculares dos gases da atmosfera são maiores que os constituintes do vapor de água.
- não depende da frequência de oscilação para um dado meio

A escolha da frequência de trabalho é uma solução de compromisso entre a resolução pretendida (que cresce linearmente com a frequência) e o alcance (que diminui com o quadrado da frequência – veja fig. 2.2). O alcance diminui em função da atenuação, que se deve à absorção, pelas partículas do meio, de parte da energia de movimento, convertendo-a em calor. A absorção não depende das perdas de energia por reflexão, difração, refração e dispersão, que só alteram a direção da propagação [53].

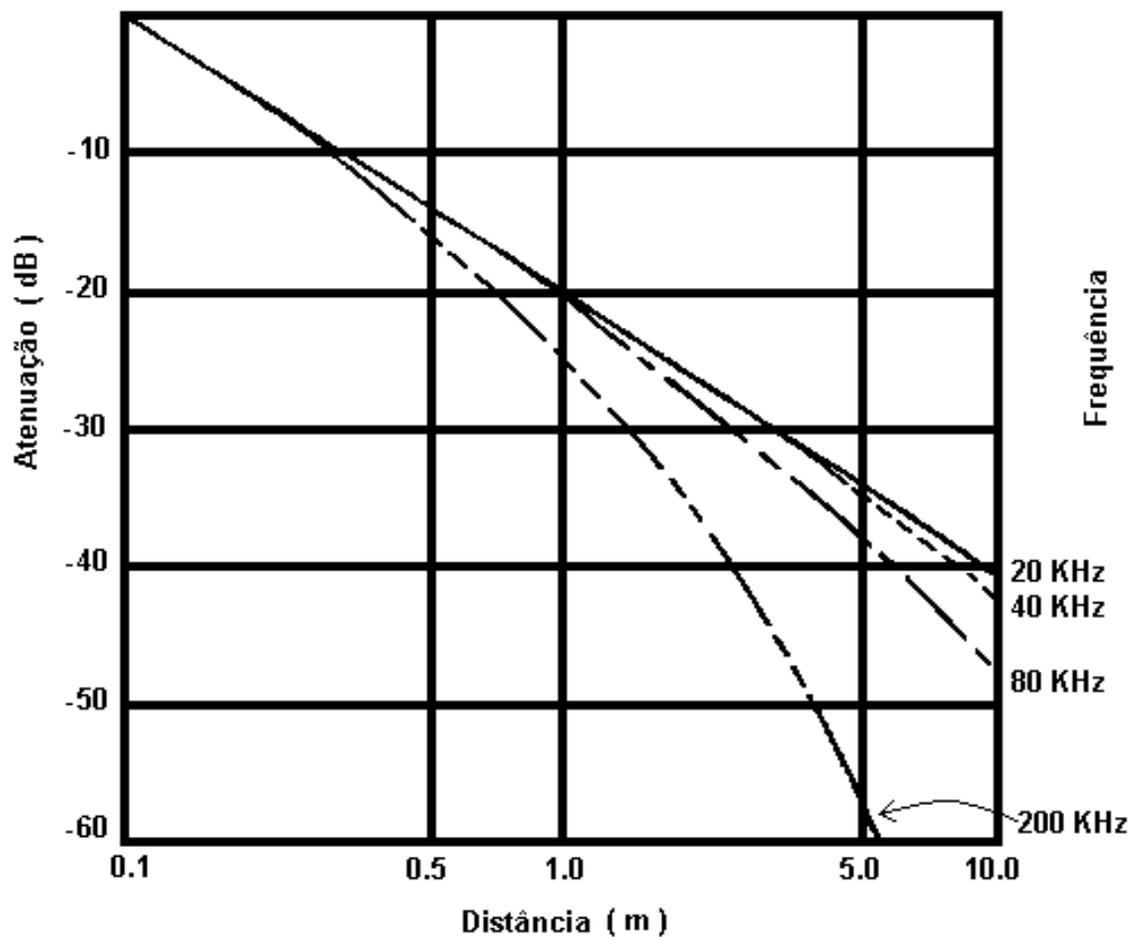


Figura 2.2 – Gráfico Atenuação x distância

Os fatores que diferenciam os sinais encontrados nos diversos sistemas de localização estão ligados ao ambiente em investigação e às propriedades que se quer conhecer do objeto a localizar. Estes fatores são determinantes das características dos subsistemas eletrônicos de transdução, condicionamento, processamento dos sinais e apresentação dos resultados. Este trabalho explora as características da propagação de ultra-som a 40 KHz no ar, em visada direta (sem reflexões), evitando assim os maiores problemas de perdas por absorção e dispersão de ondas encontrados nos sistemas baseados no método de pulso-eco. As questões referentes à escolha do transmissor e dos sensores serão vistas com detalhes mais adiante, neste mesmo capítulo.

Os sinais captados por sonares em ambiente marinho são geralmente sobrepostos com intenso ruído, não sendo incomum encontrar 30 vezes mais energia de ruídos que de sinais de interesse. Além disso, os movimentos das massas de água, da superfície do mar, dos alvos investigados e da plataforma onde se fixam os transdutores provocam uma grande variedade de distorções nas medidas de tempo de propagação dos sinais acústicos. Para que um sistema

de sonar seja efetivamente útil, deve retirar os ruídos, compensar as distorções, anular os efeitos das reflexões no fundo e da superfície do mar, e processar sinais digitalizados com algoritmos bastante complexos em tempo real [36,57]

2.2.4 Sistema de Sonar dos Morcegos

Um bom exemplo encontrado na Natureza de orientação com o uso de sinais acústicos é encontrado no sistema sensorial dos morcegos, adaptado para a caça noturna. Eles podem caçar insetos em pleno vôo graças a um sofisticado sistema de transmissão, recepção e interpretação de ondas acústicas, ainda muito avançado em comparação com o atual estado-da-arte de sistemas de SONAR feitos pelo Homem.

Os morcegos da espécie *Pteronotus parnelli rubiginosus* (morcego de bigode), quando iniciam a perseguição a um inseto, emitem um sinal acústico de alta intensidade, que se inicia com frequência constante para uso de efeito Doppler, seguido de um sinal cuja frequência diminui rapidamente (*chirp*) para medir as características de pulso-eco (tempo de propagação e amplitude). Esta segunda porção permite que sensibilidade aumente, pelo uso de processamento bastante sofisticado (compressores de pulso). Todo o sinal se caracteriza por conter harmônicos predominantemente na faixa de frequências de ultra-som (10 kHz-100kHz). Quando um inseto é localizado (pelas características dinâmicas de vôo), o morcego vai em sua direção, e à medida que se aproxima do alvo, os sinais vão se tornando mais curtos e com menores intervalos entre pulsos consecutivos, para aumentar a precisão. O efeito Doppler é usado numa forma de medida de velocidade do batimento das asas e é associado à detecção de ecos para classificação de alvos. Várias bandas de frequências sofrem separação por filtros de alto fator de qualidade e discriminadores de tipo de modulação de sinal de onda contínua. Os diferentes modos de interpretação dos sinais, e a conjugação de duas ou mais destas informações (correlatores) são feitos por uma rede de neurônios que integra várias regiões especializadas do córtex. É um exemplo biológico de sistema adaptativo baseado em redes neurais. Maiores detalhes podem ser vistos em [5] ou na literatura especializada em neurofisiologia. Muitos trabalhos de pesquisa de navegação automática de robôs tomam como ponto de partida o modelo do morcego e outros exemplos biológicos, como pode ser visto em [7,14,39].

2.2.5 Processamento Digital de Sinais

Assim como os morcegos fazem um processamento sofisticado e dinâmico dos sinais acústicos, a engenharia de sistemas de sonar/radar tem usado extensivamente as técnicas de Processamento Digital de Sinais (*Digital Signal Processing - DSP*). Estas técnicas podem ser vistas de forma intuitiva como a substituição funcional de circuitos analógicos de filtros e similares por microprocessadores especializados, que fazem tratamento matemático de sinais analógicos convertidos em digitais, em tempo real (ou seja, com velocidade igual ou maior que os filtros analógicos). O “tempo real” foi viabilizado para uso difundido (em eletrodomésticos, não mais restrito aos equipamentos militares) nas duas últimas décadas, com o barateamento de microprocessadores que, em vários aspectos, são especificamente

projetados para a implementação de algoritmos otimizados de processamento digital, já conhecidos desde a década de '60. Em uma visão mais ampla, os DSP's (como são conhecidos os *Digital Signal Processors*) são microprocessadores com estrutura otimizada (barramentos múltiplos, multiplicador paralelo, unidades aritméticas para cálculos simultâneos de endereços e dados, operação de vários processadores em paralelo) para executar operações repetitivas de multiplicar-acumular, que aceleram o cálculo da Transformada Rápida de Fourier (*Fast Fourier Transform*, ou simplesmente FFT), operação básica em uma infinidade de algoritmos para processamento de sinais. A operação de multiplicar-acumular acelera também os filtros digitais de implementação direta, as redes neurais, o processamento de imagens por convolução e os filtros espaciais para matrizes de sensores (*beamforming*).

Os sistemas de sonar dependem invariavelmente do uso de filtros adaptativos sofisticados, pois a energia do sinal de eco em ambiente marinho, na maioria das vezes, é uma ordem de grandeza menor que a do ruído. O que torna possível a detecção de sinais é sua estabilidade em termos estatísticos, enquanto os ruídos são geralmente aleatórios. Além disso, problemas graves de distorção dos sinais por reflexões múltiplas na superfície e no fundo do mar determinam o uso de técnicas de processamento digital de sinais específicas para a natureza dos sinais envolvidos e para o controle da diretividade dos sistemas de sensores, geralmente formada por muitos deles dispostos em forma de matriz .

Uma matriz de sensores é um grupo distribuído espacialmente em uma geometria específica. Se posicionarmos uma matriz de sensores em um campo de ondas de diversas fontes, técnicas de processamento digital para tratamento espacial ou espaço-temporal permitem discriminar sinais vindos de uma direção específica, evitando os ruídos e os sinais de fontes que não interessam. Esta técnica, conhecida como *beamforming*, já mencionada neste trabalho, é um processo usado para filtragem espacial, obtido geralmente pela combinação linear dos sinais dos sensores da matriz, criando uma saída única com a mencionada discriminação de ganho direcional. A conformação da matriz de sensores com espaçamento conhecido permite a amostragem de séries de valores instantâneos defasados se a onda acústica incide em ângulo diferente de 90°; o sinal de cada sensor é amostrado com um atraso programável (que determina a direção de maior ganho da matriz), passa então por um ganho específico, de forma a termos uma configuração similar à de um filtro digital (temporal), onde os defasamentos de uma amostra (z^{-1}) são substituídos pelo defasamento espacial. Desta forma, poderíamos usar as redes neurais com características dinâmicas [28,30], por exemplo a conhecida como time-delay neural network (TDNN) , pois teríamos uma estrutura de filtro bem adequada ao tipo de sinais em questão e a vantagem de ser adaptativo. Esta técnica é conhecida como filtragem espacial, por determinar um maior ganho do conjunto-sensor em uma direção específica.

Os pesos individuais de cada sensor formam um padrão de ganhos dependentes da direção em que chega o sinal, de forma a permitir a mudança da direção de maior ganho da matriz de sensores pela alteração criteriosa dos ganhos dos sensores. Da mesma forma, podem ser criadas direções de grande atenuação evitando os sinais recebidos de uma fonte em específico, o que chamamos de filtro espacial.

Em sistemas de SONAR, estas são práticas difundidas, e cumprem bem seu papel também em outras classes de problemas com características semelhantes na conformação espacial de seus sinais ou com demanda de grande desempenho em relação à precisão e à densidade de detalhes (formação de imagens). Outros exemplos são os equipamentos de radar, imagens médicas, rádio-telescópios matriciais, observação de atividade sísmica (incluindo a prospecção de petróleo), etc. [11,24,27,36,38,54,65].

Por meio de *beamforming*, isoladamente, não se consegue obter a distância absoluta do objeto, mas sim através da conjugação com outras técnicas. Várias conjugações das técnicas mencionadas se tornam apropriadas em aplicações onde se procura obter maior sensibilidade a algum aspecto do objeto e maior rejeição das interferências do ambiente. Nos sistemas de SONAR para investigação no mar, por exemplo, redes neurais de alto desempenho têm sido usadas para predição do movimento das ondas, que tem grande parcela no modelo matemático de distorções de sinais naquele ambiente. O fato das ondas do mar se caracterizarem por ter movimento caótico (cuja descrição pode ser aproximada por funções caóticas, de caráter determinístico) torna bem adequado o uso de preditores com redes neurais [15,29,30,34,57].

Da mesma forma que é criado um padrão direcional de ganhos na recepção, também se pode implementar um controle de defasamentos programáveis para os sinais a serem enviados a uma matriz de transmissores, de forma a criar uma frente de ondas que se propaga em uma direção específica, processo conhecido por “*inverse beamforming*”. [36]

O posicionamento e a sensibilidade dos sensores têm geralmente uma confiabilidade limitada e instável, sugerindo a parametrização do modelo do sistema especificamente para o ambiente em investigação, gerando dificuldades adicionais na calibração dos sensores. Esta calibração tem sido alvo da maior atenção por parte dos pesquisadores dedicados ao progresso dos sistemas de SONAR e de diagnóstico médico por ultra-som [12,19,42,58] . Nestes últimos, o método de medição usado é de pulso-eco, com as dificuldades já mencionadas. O número de sensores é grande, sendo comum vermos 64, 128, 16 X 16 (painéis com matrizes em 2 dimensões) , tornando essencial a adaptabilidade dos parâmetros dependentes da posição dos sensores.

2.3 Método de Medição de Posição Proposto

Neste trabalho é proposto um sistema para medição de posição, em três dimensões, de objetos que possam conter um transmissor. Compõe-se de um pequeno transmissor de ultra-som e de infra-vermelho, um conjunto de sensores e um sistema de aquisição de intervalos de tempo por computador. É usada uma conjugação de transmissão de ondas pulsadas e a recepção com detecção de limiar de envoltória. A propagação de ondas diretamente de transmissor para receptor proporciona baixíssimos níveis de ruído e distorções, de forma que as preocupações quanto à relação sinal/ruído são suficientemente atenuadas com um estágio simples de condicionamento de sinais. Um dos parâmetros mais fortes de projeto foi a simplicidade de implementação, com uso de componentes comuns, de baixo custo. Pode ser

viabilizado assim o seu uso em computadores pessoais ou em qualquer sistema que faça captura de movimentos de objetos ou de suas partes, aonde puder ser anexado o transmissor.

O método proposto começa pela medição dos tempos de propagação das ondas de ultra-som entre o transmissor e cada um dos 3 sensores empregados (veja item 2.4 - Sistema de Transmissor-Sensores). Um ciclo de medição de tempo se inicia com a transmissão de um trem de pulsos (8 pulsos) de onda de ultra-som junto com um sinal idêntico (em frequência e fase) de infra-vermelho. A onda de luz modulada chega a um sensor de infra-vermelho do módulo analógico (bem próximo dos sensores de ultra-som) após um tempo desprezível, se comparado com o tempo de propagação do ultra-som. Isso se explica por ser a velocidade da luz muito maior que a do som: 300.000.000 m/s para a luz e 340 m/s para o som (900.000 vezes maior). A borda de subida do 1º pulso (dos 8) do sinal de infra-vermelho é usado então para marcar o instante de início de propagação da onda de ultra-som, e portanto inicia a contagem de tempo (pulsos) dos três temporizadores (contadores digitais). A onda sonora, após alguns milissegundos, alcança cada um dos sensores em tempos quase sempre diferentes (podem ser iguais apenas nos pontos pertencentes ao lugar geométrico dos pontos equidistantes aos três sensores, uma reta, que no caso do triângulo equilátero passa pelo centro do mesmo, ortogonal ao plano que ele gera). A contagem é interrompida em cada contador quando a borda de subida do 1º pulso do sinal de ultra-som vindo do transmissor é recebida por cada um dos sensores.

O maior problema encontrado na medida de distâncias por este método consiste em determinar o instante de chegada do sinal de 40 KHz na recepção, e provoca erros significativos, mas com valores muito próximos para os três, permitindo assim serem subtraídos se necessário. Neste trabalho não se fez necessário este ajuste, pois os efeitos destes erros de medição são minimizados com a normalização feita nos na seqüência de pontos de um gesto. Existem métodos matemáticos sofisticados para correção destes erros absolutos; o uso de interpolação parabólica e uma comparação com outros mais tradicionais, podem ser vistos em [7].

O trem de pulsos gerado tem 8 pulsos na frequência de 40 KHz, e após estes pulsos o transmissor fica inativo por um tempo equivalente a 504 pulsos, para garantir que os receptores recebam o sinal antes que seja transmitido outro, estabelecendo assim um limite superior para a distância de operação em 4,35 m (veja item 2.4 - Sistema de Transmissor-Sensores). Os limites impostos por outros parâmetros dos circuitos usados são menos significativos, resultando em um limite prático de distância de aproximadamente 3 metros.

Os valores dos 3 tempos assim obtidos, são enviados a um computador que inicia o processamento destes sinais multiplicando-os pela velocidade de deslocamento de ondas acústicas no ar, para obter diretamente os valores das distâncias do transmissor até cada um dos três sensores (veja item 2.5).

De posse destas distâncias, é feito um processo semelhante ao conhecido como “triangulação”, empregado para localizar transmissores de rádio em um mapa bidimensional, usando receptores direcionais para avaliar dois ângulos, associados a dois vértices (de posições arbitrárias) da base de um triângulo e calcular a posição do terceiro vértice.

Para fazer uma comparação: o processo aqui proposto, se limitado a duas dimensões, avaliaria as distâncias absolutas entre o transmissor e mais de um receptor de posições conhecidas e, usando relações trigonométricas (interseção entre duas circunferências, por exemplo) , conseguiria avaliar a direção da fonte de sinal.

Se houvesse o limite de trabalho em duas dimensões (empregando apenas dois sensores) , teríamos apenas como avaliar a distância até o ponto de origem do sistema de coordenadas em uso e um ângulo de direção com relação ao mesmo sistema de coordenadas (deslocamento lateral), mas não seria possível diferenciar seu ângulo de elevação. Poderíamos avaliar a posição do transmissor se houvesse a garantia de que ele estivesse sempre na mesma altura que os sensores (valor da componente de uma terceira dimensão sempre igual a zero) , que não é o caso aqui tratado. Por outro lado, se forem usados mais que três sensores haveria redundância de informações, pois não há dois pontos distintos com o mesmo trio de distâncias entre transmissor e receptor.

Aqui se procura a localização em três dimensões, portanto, para acrescentar a informação da terceira coordenada, é necessário usar pelo menos três sensores não coplanares com o transmissor. Foi adotada uma configuração que se compõe de um transmissor e três sensores dispostos em um triângulo equilátero (veja figura abaixo).

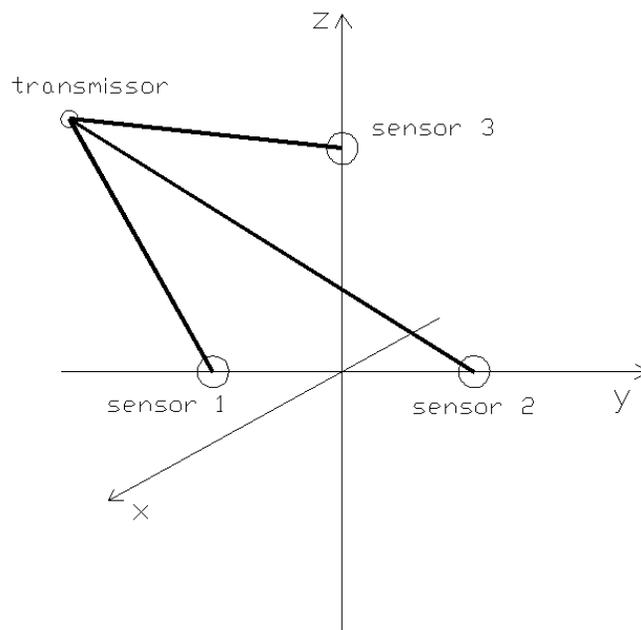


Figura 2.3 – 3 sensores

O cálculo das coordenadas x,y,z de um ponto a partir das distâncias do transmissor até cada um dos três sensores é feito com o uso de um sistema de três equações simultâneas com

três variáveis. São referentes a três esferas, e a solução deste sistema de equações identifica sua interseção, em dois pontos, um em cada um dos semi-espacos gerados pelo plano dos sensores. Só interessa o ponto em frente dos sensores, sendo desprezado o ponto com coordenada (x) negativa. Destas esferas se conhece seus centros nas posições dos sensores, se sabe que um ponto de suas superfícies coincide com a posição do transmissor, e se conhece seus raios, iguais às distâncias dos sensores – centros – ao transmissor. Isto se deriva da definição de uma esfera, como o lugar geométrico dos pontos equidistantes de um ponto, seu centro. Se tomarmos a posição da origem do sistema de coordenadas retangulares $(0,0,0)$ relativa à posição dos sensores, teremos então os valores das coordenadas dos sensores, na versão final, como na figura 2.4:

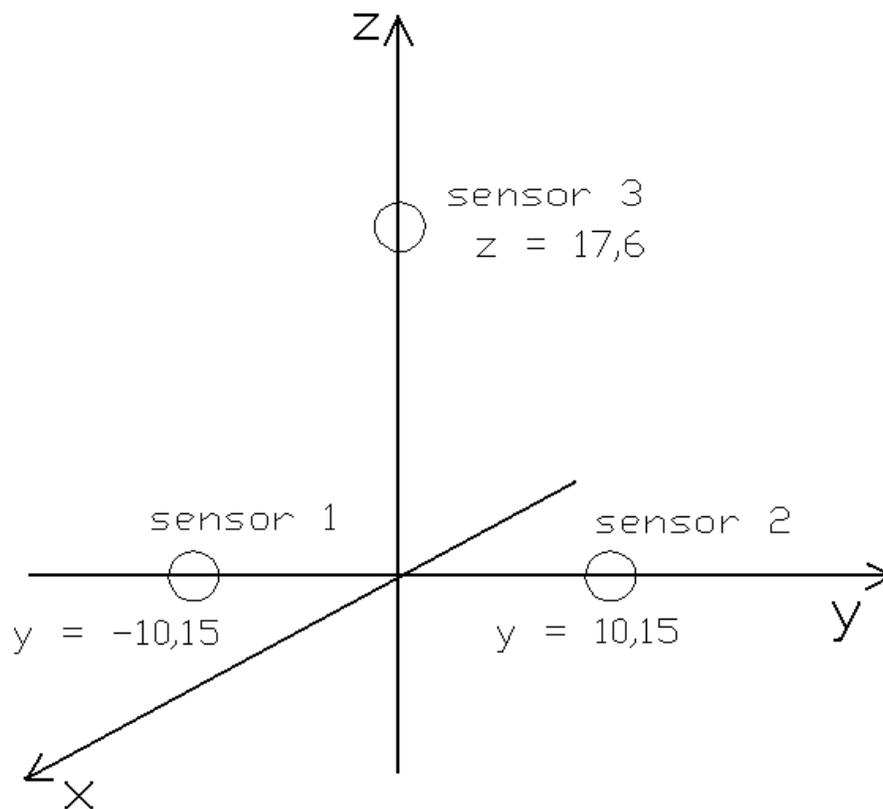


Figura 2.4 – Coordenadas dos sensores

Com o que conhecemos das esferas em questão podemos montar as seguintes equações, onde as incógnitas serão as coordenadas x,y,z do transmissor, resultantes imediatas da solução do seguinte sistema de equações:

$$\left\{ \begin{array}{l} (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 = R_1^2 \\ (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 = R_2^2 \\ (x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 = R_3^2 \end{array} \right.$$

$$\left\{ \begin{array}{l} (x - 0)^2 + (y + 10,15)^2 + (z - 0)^2 = R_1^2 \\ (x - 0)^2 + (y - 0)^2 + (z - 17,6)^2 = R_2^2 \\ (x - 0)^2 + (y - 10,15)^2 + (z - 0)^2 = R_3^2 \end{array} \right.$$

A solução deste sistema, obtida por substituições, pode ser toda expressa em função das distâncias conhecidas (dados de entrada), ou usando os valores calculados progressivamente para não cair em repetição de cálculos já executados (conveniente em um algoritmo para cálculo por computador). O resultado é:

$$y = \frac{(R_1^2 - R_2^2)}{40,6}$$

$$z = \frac{5,86 + (R_2^2 - R_3^2 + 20,3 \times y)}{35,16}$$

$$x = \sqrt{R_1^2 - 10302 - y^2 - 20,3 \times y - z^2}$$

Estes cálculos são feitos após o armazenamento pelo computador das amostras de tempos de um gesto completo, conversão de tempos em distâncias, e após um processo de filtragem descrito mais adiante.

O uso de quatro sensores poderia ser vantajoso se fossem posicionados de forma a tornar as medidas de coordenadas de três sensores relativas ao quarto sensor, por exemplo posicionando-o na origem de um eixo coordenado e os outros três nos eixos, como na fig. 2.5. Com este procedimento se consegue uma aceleração nos cálculos, pois as distâncias até o sensor localizado na origem são obtidas a partir das coordenadas em questão menos zero, além do que uma das coordenadas de cada um dos outros sensores é igual a zero, economizando algumas subtrações.

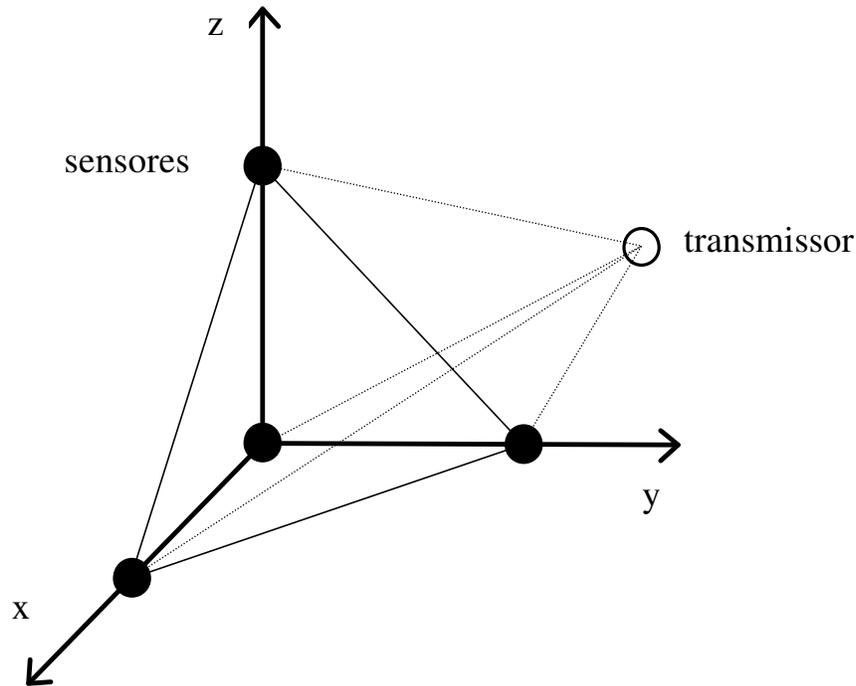


Figura 2.5 – Com 4 sensores

O circuito eletrônico do protótipo está preparado para comportar o 4º sensor, pois se houver problemas com tempo de cálculo das coordenadas em tempo real em alguma aplicação futura, basta usar a simplificação de cálculo mencionada.

Para adequar a aquisição de pontos à composição de gestos e atendendo ao critério de simplificação do problema, foi ignorada a coordenada x, associada à noção de profundidade, pouco significativa se nos limitarmos ao desenho de letras no espaço. A figura 2.6 mostra um exemplo com a letra C que ilustra este fato (escalas em cm).

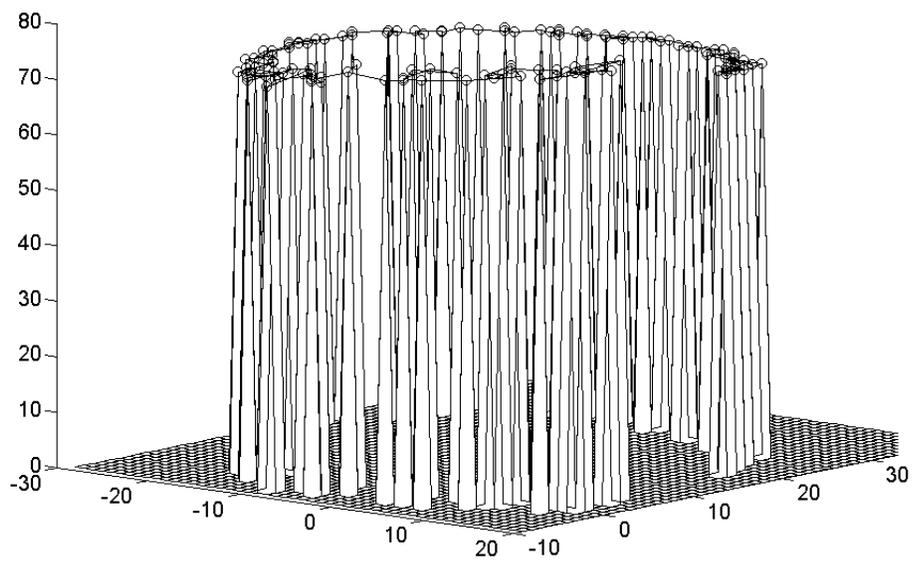
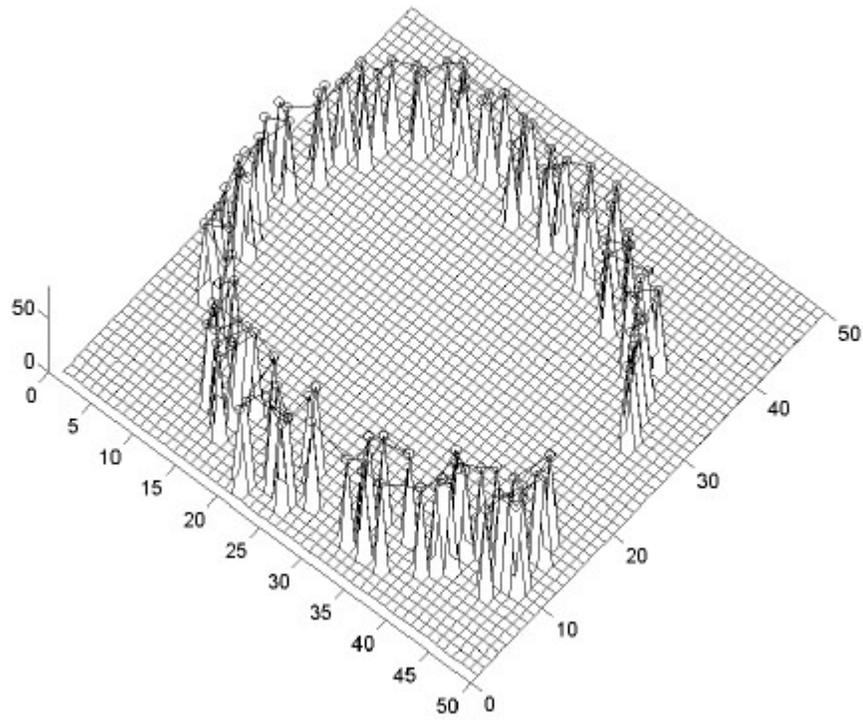


Figura 2.6 – Representação em 3 dimensões

Inicialmente, foi tentada a disposição dos sensores nos vértices de um triângulo de 7,0 cm de lado (fig. 2.7), soldados diretamente na placa de circuito impresso, para primeiras avaliações. Foram feitos alguns testes, gerados gráficos com o resultado da aquisição de alguns gestos, nos quais se notavam ruídos de amplitudes consideráveis superpostos ao que se esperava como resposta aos gestos.

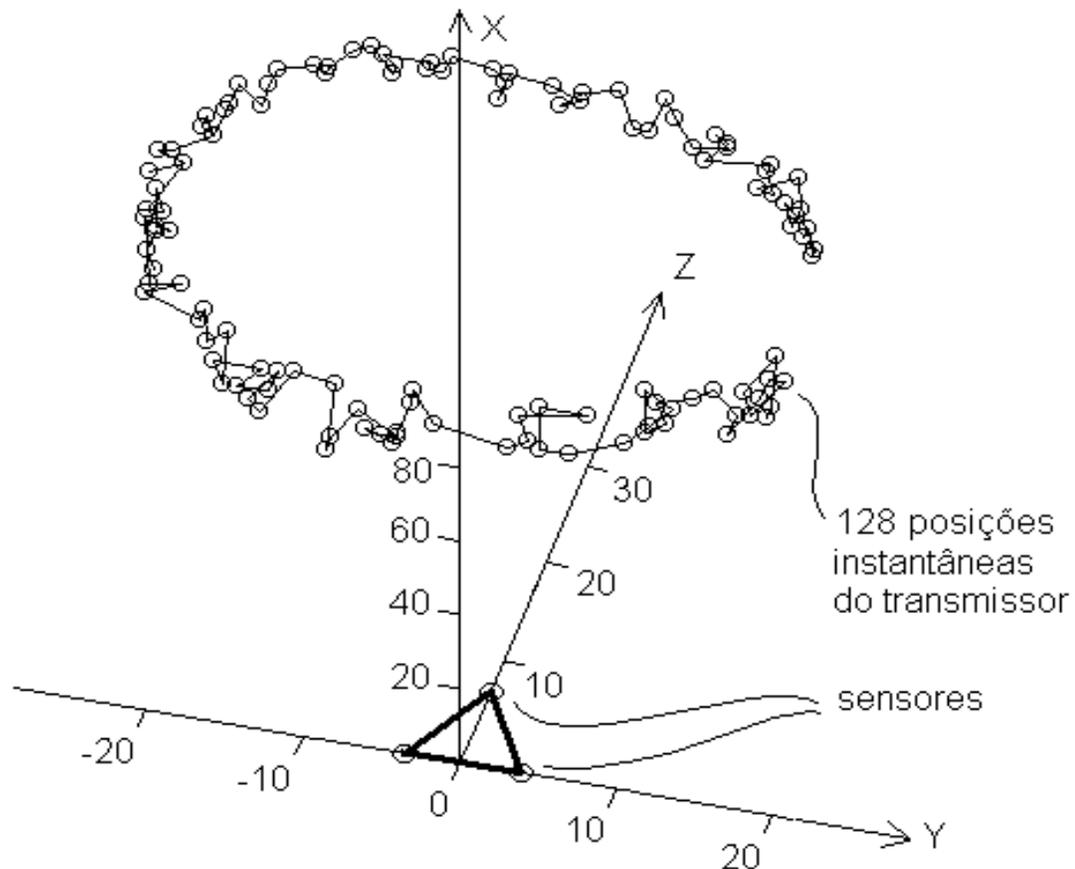


Figura 2.7 – Posição relativa entre os sensores e um gesto

Foi acrescentado um processo de filtragem, por média simples entre dois pontos adjacentes, dos seus valores de distância até o transmissor, para suavizar o percurso do gesto, mas os resultados, apesar de apresentarem alguma melhoria, não foram muito satisfatórios.

A figura 2.8 (escalas em cm) mostra gestos que desenhavam as letras C e V (dois exemplos para cada letra), com a distância entre sensores de 7 cm, comparando os resultados obtidos antes da implementação do filtro de média (coluna da esquerda) e depois da implementação do mesmo (coluna da direita). Nota-se um ganho razoável em qualidade, mas ainda muito desfavorável para o reconhecimento por uma rede neural, pela presença forte de ruídos.

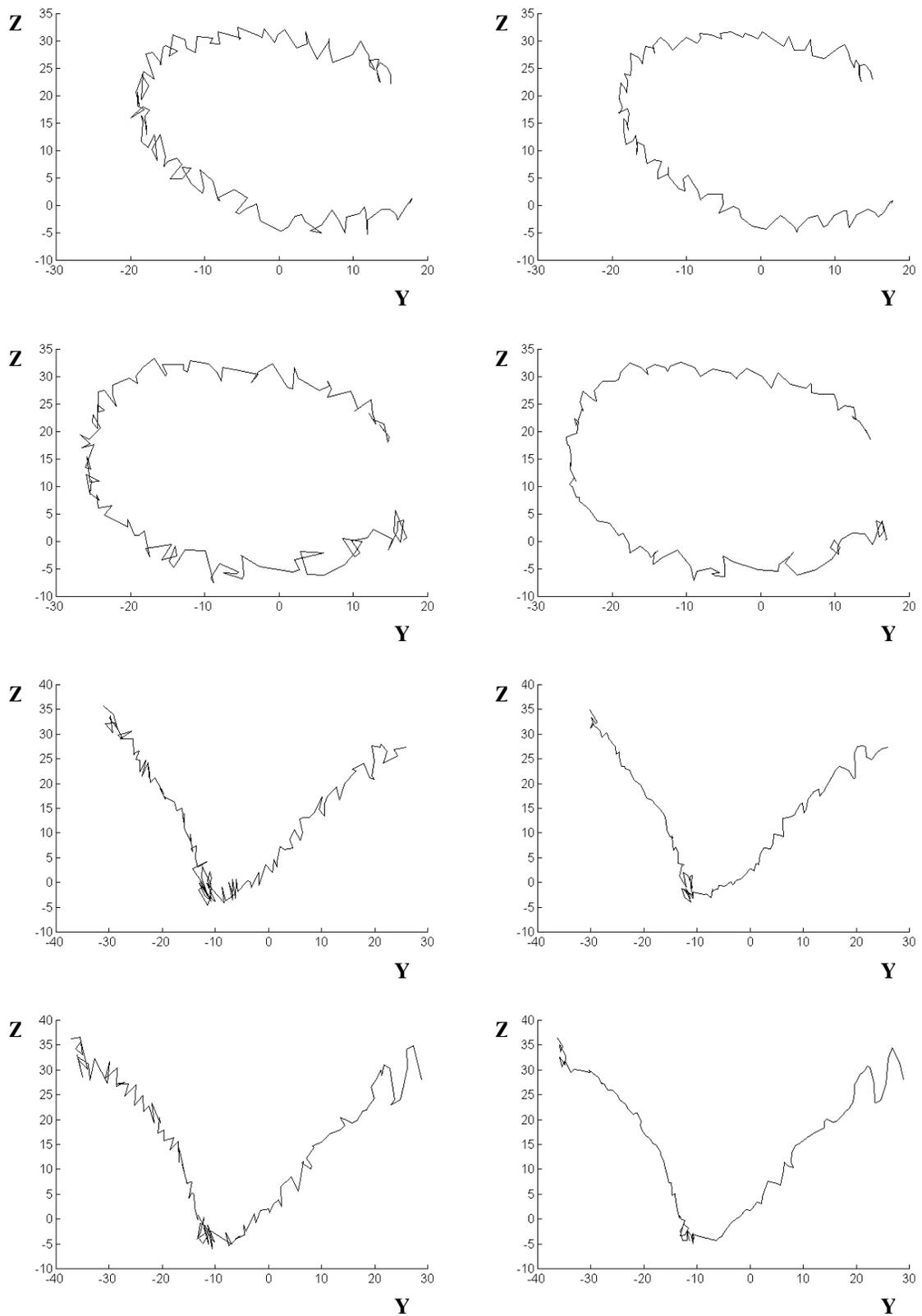


Figura 2.8 – Gestos amostrados com sensores afastados de 7 cm

A partir do estudo de dificuldades de convergência de redes neurais na tentativa de solução de um problema bem semelhante, já havia um conhecimento prévio de que poderia surgir uma dificuldade de cálculo das coordenadas a partir das três distâncias medidas. Esta dificuldade seria gerada pelas grandes diferenças de escala entre os valores absolutos das distâncias medidas (medidas em metros) e, as pequenas diferenças entre elas (medidas em milímetros) que permitem obter as distâncias calculadas. Para melhor visualizar esta relação de escalas de distâncias, podemos supor uma montagem com apenas dois sensores, e comparar a distância sensor-transmissor (metros) com a distância percorrida pela onda entre os instantes em que alcança os diferentes sensores (mm). Veja figura 2.9.

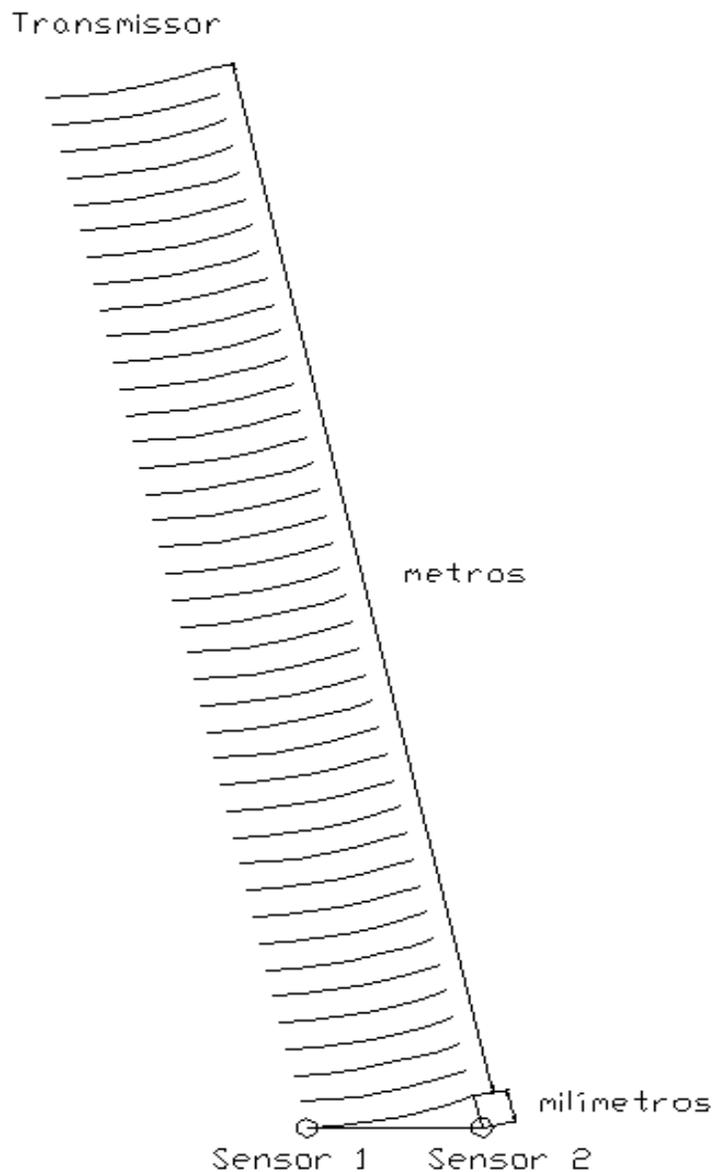


Figura 2.9 – Relação entre escalas de distâncias

Fica claro que as variações angulares na posição do transmissor provocam pequenas mudanças na diferença de tempo entre a chegada da onda esférica ao sensor mais próximo e a chegada ao mais distante. Desta forma, erros considerados pequenos na avaliação da distância transmissor-sensor terão influência significativamente maior na diferença de tempo de propagação, que é o que determina a diferenciação de pontos no espaço, de forma não explícita, na solução da interseção entre as esferas em coordenadas retangulares (fica mais explícito usando coordenadas polares). Os mencionados erros considerados pequenos podem ter origem na resolução dos contadores de tempo, nos ruídos inerentes aos componentes eletrônicos associados aos receptores, e principalmente no limite físico da resolução da medida com uso de ultra-som a 40 KHz (veja mais adiante, item 2.5).

Diante destas considerações, se percebe que aumentar a separação entre os sensores diminui os efeitos dos erros de medição de tempo sobre o cálculo da posição em coordenadas retangulares, pois fica mais pronunciada a diferença entre os tempos de chegada da onda, e maior também a sua variação para cada 1° (um grau) de variação angular da posição do transmissor. Aumentando esta diferença entre tempos, melhora a capacidade de diferenciação de pontos no espaço.

A distância entre sensores foi então aumentada para 20,3 cm. A escolha por esta medida se deu por uma simples conveniência, por ser esta uma das dimensões da placa de circuito impresso do módulo analógico, permitindo então a fixação dos sensores sem o uso de hastes de extensão, pouco estáveis. Poderiam ser testadas distâncias intermediárias ou maiores, para estabelecer uma relação deste afastamento com os níveis de ruído, mas isto seria uma busca de otimização que excede os objetivos deste trabalho. Só seria ultrapassado o limite da placa se fosse necessário, sendo conveniente evitar que o tamanho do conjunto se torne um fator de incômodo no seu uso.

O ganho de qualidade da medida com o afastamento dos sensores foi grande, como se pode perceber nas figuras da página seguinte, gestos que desenhavam as letras C e Z, coluna da esquerda, com a distância entre sensores aumentada para 20,3 cm e já com o uso do filtro de média (escalas em décimos de milímetros, usada para eliminar quaisquer preocupações com erros de arredondamento).

Para tornar o formato os dados mais adequados para as redes neurais usadas no reconhecimento de gestos, foi necessário fazer a reamostragem das seqüências de pontos para uniformizar os deslocamentos e a normalização dos comprimentos dos segmentos que ligam dois pontos consecutivos, resultando nas figuras da página seguinte, coluna da direita, nas mesmas condições da coluna da esquerda, para comparação dos formatos dos gestos e dos níveis de ruído. As escalas destes gráficos não têm mais correspondência com as dimensões reais do gesto, pois a normalização torna a distância entre dois pontos quaisquer igual a 1.

O processo de reamostragem com deslocamentos uniformes e normalização será bem detalhado no capítulo 3, item 3.3, já tendo sido esclarecidos os aspectos relevantes das redes neurais que são favorecidos e a sua relação com este processo de pré-processamento.

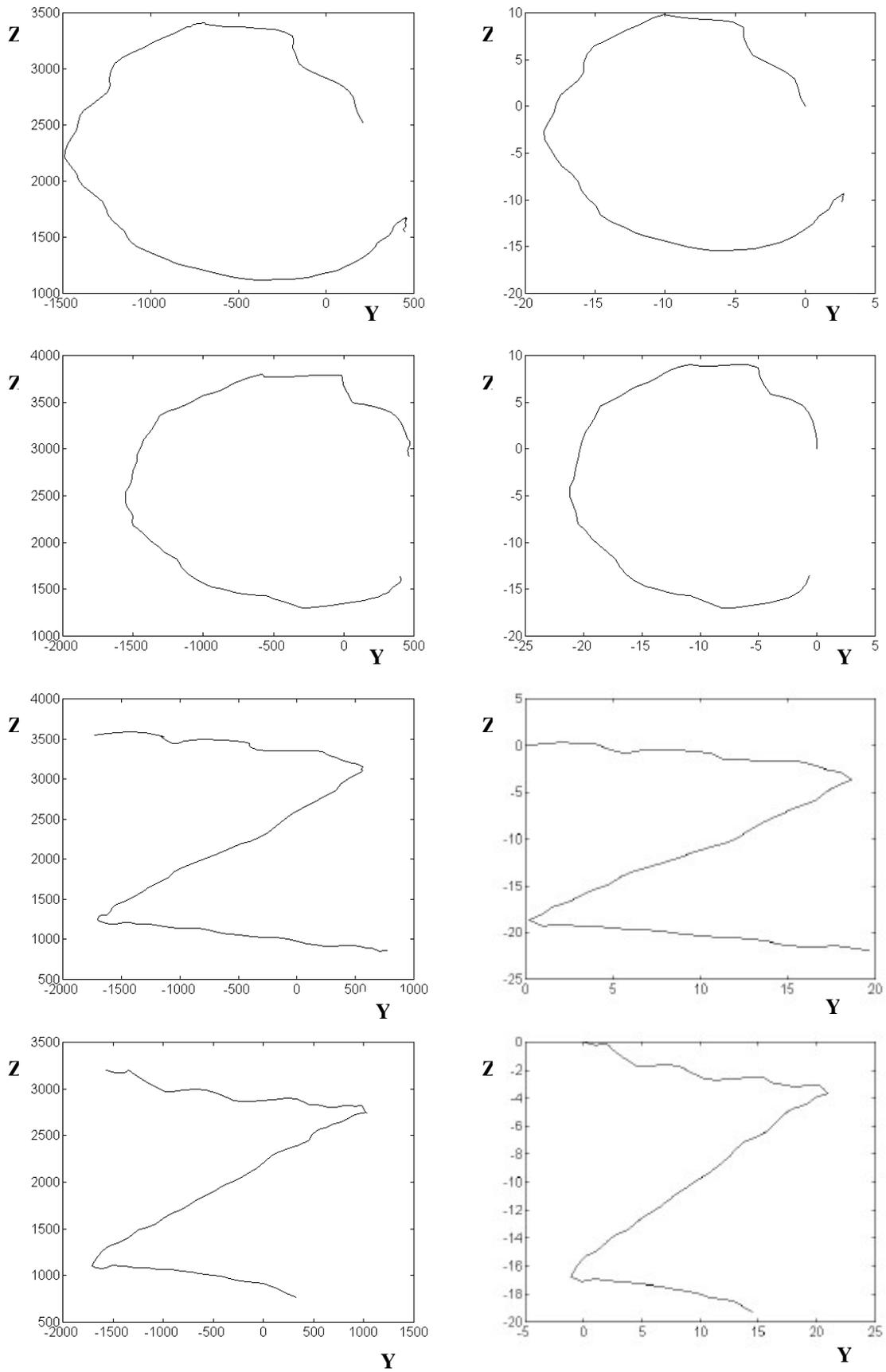


Figura 2.10 – Gestos amostrados com sensores afastados de 20,3 cm

Foram feitas algumas experiências com outro modelo de par transmissor-receptor (transdutores) do mesmo fabricante, de mesma frequência, mas com maior diâmetro, de diretividade mais acentuada. Não houve ganho algum na qualidade da medida, alterando-se apenas a diretividade e o alcance, como era esperado. Disto se conclui que as ondulações que restam perceptíveis nos gráficos dos gestos se devem à resolução intrínseca da medida, relacionada apenas à frequência de operação, adicionada a um remanescente de erro gerado pelas diferenças de escala, com solução pelo afastamento dos sensores entre si.

Tendo sido definidos os aspectos de montagem dos transdutores, pôde ser iniciada a aquisição de gestos para treinamento e avaliação de uma rede neural para reconhecimento dos mesmos. Esta aplicação foi escolhida por oferecer alguns desafios em termos de formatação de seqüências de amostras e por ser uma forma bem abrangente de comprovar a validade dos métodos de medição e aquisição de dados empregados.

2.4 Sistema de Transmissor-Sensores

Com o propósito de comprovação prática do método de medição idealizado, foi projetado e construído um protótipo de circuito eletrônico, e foram desenvolvidos programas para computadores pessoais para aquisição e interpretação de dados. Neste tópico serão descritos os circuitos eletrônicos no estágio final do protótipo, e detalhados alguns deles pela relevância que tenham no funcionamento e pela importância de se mencionar a sua evolução a partir das primeiras idéias.

O sistema eletrônico desenvolvido pode ser dividido em três blocos: transmissor, conjunto de receptores (também referenciado como módulo analógico), e módulo digital, separados em três placas de circuito impresso (fig. 2.11). Um computador pode ser incluído como um 4º bloco, porque processa e reage ao que é captado pelos sensores.

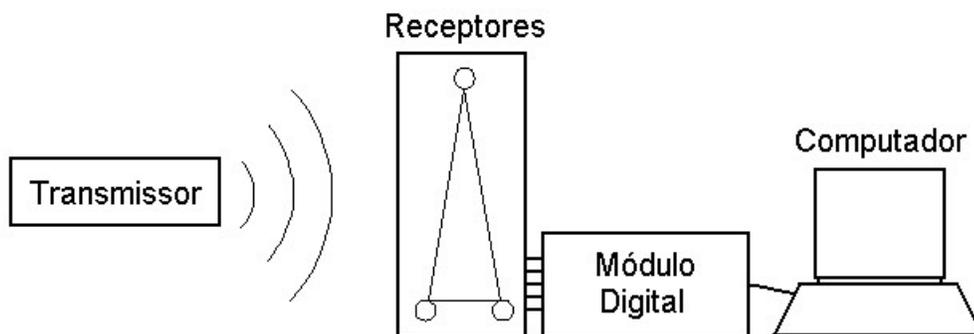


Figura 2.11 – Diagrama de blocos geral

As medidas dos tempos de propagação do ultra-som para os três sensores e seu processamento para obtenção das coordenadas x,y,z, são feitos em tempo real, a uma taxa de quase 80 pontos por segundo. Esta cadência está atrelada ao sinal do transmissor, pois é este que gera o sinal de infra-vermelho, que faz uma função de sincronismo, determinando o instante em que os contadores de tempo de propagação de ultra-som são reinicializados e começam a contar pulsos.

O transmissor é um módulo bastante compacto, com o propósito de ser fixado ao dedo indicador do usuário, e alimentado por bateria. Ele pode ser acionado por uma micro-chave apenas no momento de interagir com a máquina por meio de movimentos úteis. Contém um oscilador (que gera onda quadrada), um modulador para formatar os trens de pulsos e o tempo de inatividade entre dois trens de pulsos, e amplificadores de sinais para emissores de ultra-som e infra-vermelho.

A frequência de onda acústica usada é de 40kHz, localizada então no que se convencionou chamar de faixa de ultra-som. A frequência escolhida é conveniente ao usar transdutores para propagação no ar, pois tem um bom equilíbrio entre as perdas (que determinam o alcance) e a resolução. Os transdutores de frequência mais alta (200 ~ 400 KHz) , devido ao menor comprimento de onda ($\lambda=c/f$) alcançam maior resolução nas medições de tempo de propagação, resultando em uma confiabilidade na medida na ordem de décimos de milímetros, mas com um alcance de poucos centímetros, devido à atenuação elevada da energia sonora (proporcional ao quadrado da frequência da onda). Transdutores de frequências menores que a escolhida, por terem comprimento de onda maiores, comprometeriam a resolução da medida, que está em seu limite mínimo para se adequar a este trabalho (<1cm). Além disso, o aumento no alcance oferecido por estes é desnecessário, sendo satisfatório o obtido com 40 KHz [1,53]

Embora existam vários tipos de transdutores de ultra-som (capacitivos, piezoelétricos, magnetoestrictivos), poucos são adequados para a aplicação em questão. A determinação de se fazer um sistema de dimensões reduzidas e de baixo custo já limita a quantidade de tipos e modelos a escolher. Os transdutores capacitivos, de melhor acoplamento acústico com o ar [1], exigem o uso de gerador de tensão razoavelmente alta para sua polarização, complicando sua utilização em um circuito para se acoplar à mão do usuário, alimentado por bateria (que é o caso do transmissor idealizado). Segundo o que se vê em [53] , os transdutores piezoelétricos apresentam maior robustez, maior imunidade a ruídos elétricos e acústicos, menores preços, maiores amplitudes na recepção e uma relação sinal/ruído melhor, em comparação com os capacitivos. Entre os transdutores disponíveis, optou-se então pelos piezoelétricos de 40 Khz, do fabricante Murata, modelo MA40S2R, de tamanho reduzido e fácil obtenção (são usados em alarmes automotivos). São adequados em termos de sua frequência de operação, pois permitem um alcance de até alguns metros e resolução suficiente para amostrar o formato de pequenos gestos (de alguns centímetros). Têm também um ângulo de abertura na região sensível (lobo frontal) bem amplo, permitindo maior liberdade nos movimentos do usuário. Foi testado também o modelo MA40A5, do mesmo fabricante, e igualmente barato e de fácil obtenção, com a mesma frequência de operação, mas com o diâmetro 2,5 vezes maior. É um modelo mais diretivo e um pouco mais sensível, mas foi feita a opção pelo outro, menor e com ângulo de abertura maior [53]

O sinal de infra-vermelho é gerado por dois leds em série (dois, para alcançar uma potência razoável), que emitem luz com pico de energia irradiada na faixa de 950 nm, modulado pelo mesmo sinal de ultra-som de 40 KHz, em forma de trens de pulsos. Seu gasto de potência é muito maior que o do emissor de ultra-som, com picos de corrente que chegam próximo de 500 mA.

O trem de pulsos gerado é formado por 8 pulsos na frequência de 40 KHz (período de $T = 25 \mu\text{s}$), e após estes pulsos o transmissor fica inativo por um tempo equivalente a 504 períodos de mesma duração totalizando 512 pulsos de $25 \mu\text{s} = 12,8 \text{ ms}$, como na figura 2.12 :

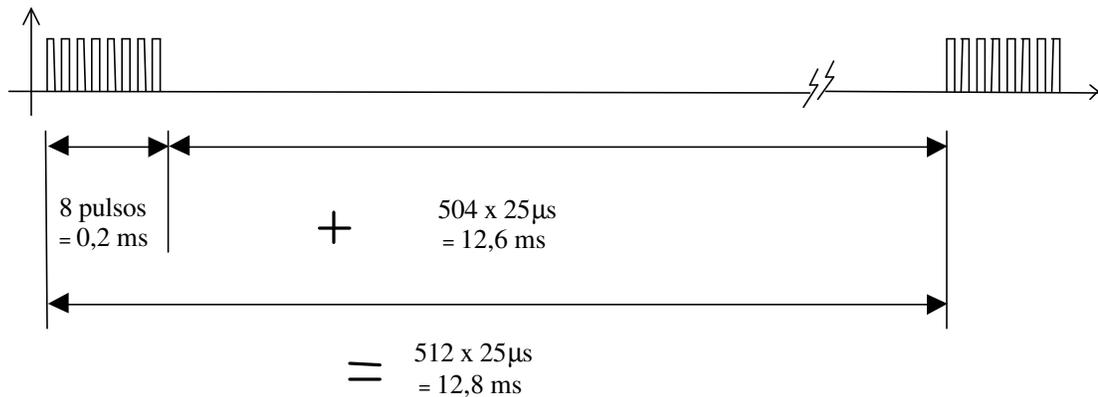


Figura 2.12 – Trem de pulsos nos terminais do transmissor

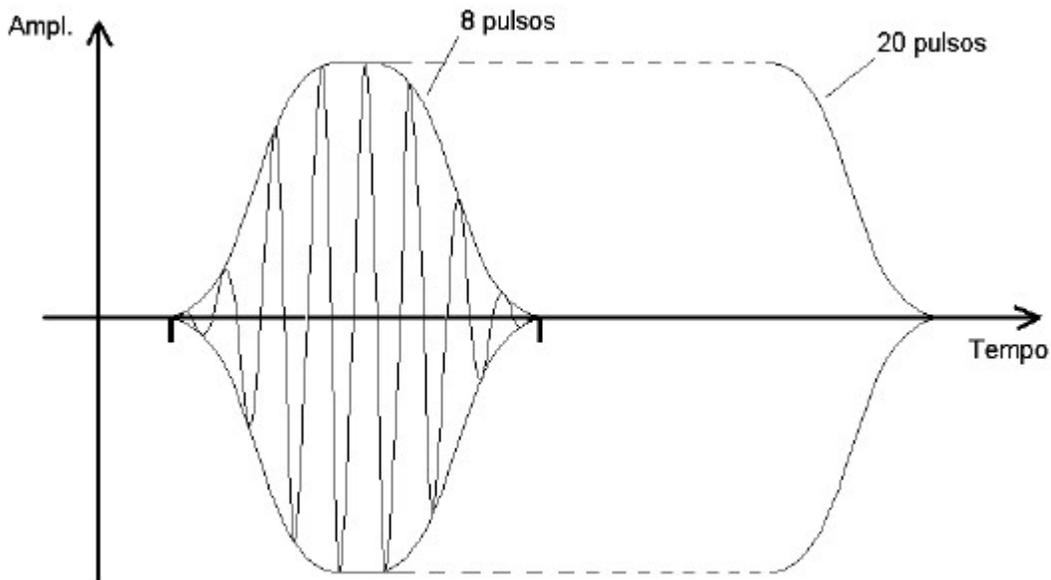


Figura 2.13 – Trem de pulsos nos terminais do sensor

O dimensionamento para o número de pulsos em 8 foi feito de maneira empírica, tendo sido experimentados de 1 a 20 pulsos e observada a envoltória do sinal recebido, que atinge seu máximo entre 5 e 6 pulsos, em média. Obedecendo ao critério de economia de bateria, o transmissor não deve ficar muito tempo neste máximo. Para aumentar o número de pulsos com o mesmo circuito do transmissor, o próximo valor possível seria de 16 pulsos, o que dobraria o consumo de energia, sem muito ganho em qualidade na envoltória do trem de pulsos recebido. Esta envoltória se deve à acomodação mecânica do transdutor até atingir a ressonância; sua descrição matemática com aproximações pode ser vista em [7]. A figura 2.13 mostra a onda de 40 KHz modulada, obtida após o receptor. O formato senoidal dos pulsos (e não mais retangular) se deve à drástica atenuação das harmônicas superiores pelos transdutores, pois respondem numa faixa estreita em torno da frequência central (40 KHz).

[figura - onda de 40 KHz obtida após o receptor]

Transmitindo um sinal neste formato, obtenho 1 amostra de tempo de propagação a cada 12,8 ms, correspondendo a uma taxa de $1 / 0,0128 = 78,125$ amostras/s. Nos gestos que foram amostrados para teste de reconhecimento por redes neurais, são feitas 128 amostras por gesto, então um gesto tem o tempo de execução fixado em $128 \times 12,8 \text{ ms} = 1,64$ segundos

A separação entre os trens de pulsos (o transmissor fica inativo por um tempo equivalente a 504 períodos) tem por objetivo garantir que todos os receptores tenham recebido um sinal de ultrassom (como o da figura anterior) antes que seja transmitido o próximo trem de pulsos (o sinal concomitante de infravermelho reiniciaria instantaneamente os contadores de tempo de propagação).

A onda sonora se propaga a 340 m/s; se o tempo entre dois trens de pulsos é de 12,8 ms, a distância máxima em metros (m) que a onda pode se propagar neste intervalo pode ser obtida multiplicando-se o tempo em segundos (s) pela velocidade em metros por segundo : $0,0128\text{s} \times 340\text{m/s} = 4,35 \text{ m}$. Este seria a restrição ao limite superior de distância para operação, feita pelo formato da onda transmitida (total de 512 pulsos). Outros limites são impostos por parâmetros do circuito usado, como a potência máxima do transmissor e o balanço entre a sensibilidade e a relação sinal/ruído na recepção, resultando em um limite prático de distância de aproximadamente 3 metros. Acima desta distância, nos gestos mais amplos, começam a acontecer falhas no sinal de recepção, ficando alguns trechos sem amostras de pontos. A característica de diretividade dos transdutores contribui para isto (atenuação maior para maiores ângulos, somada à atenuação por absorção e dispersão que já é grande a esta distância), de forma que um gesto feito a 4m com o cuidado de se manter o transmissor apontado estritamente na direção dos receptores pode ser bem amostrado. O excesso com os 4,35 m foi mantido para não aumentar o número de componentes do circuito do transmissor (mantendo os 512 pulsos – que é uma potência de dois - a decodificação do fim de um ciclo fica simples). Uma outra conveniência de se manter o tempo inativo por maior tempo vem da minimização do gasto médio de potência, principalmente do transmissor de infra-vermelho, para viabilizar a alimentação por bateria. Veja na figura 2.14 o diagrama eletrônico do transmissor, e observe o decodificador simples com diodos e um resistor.

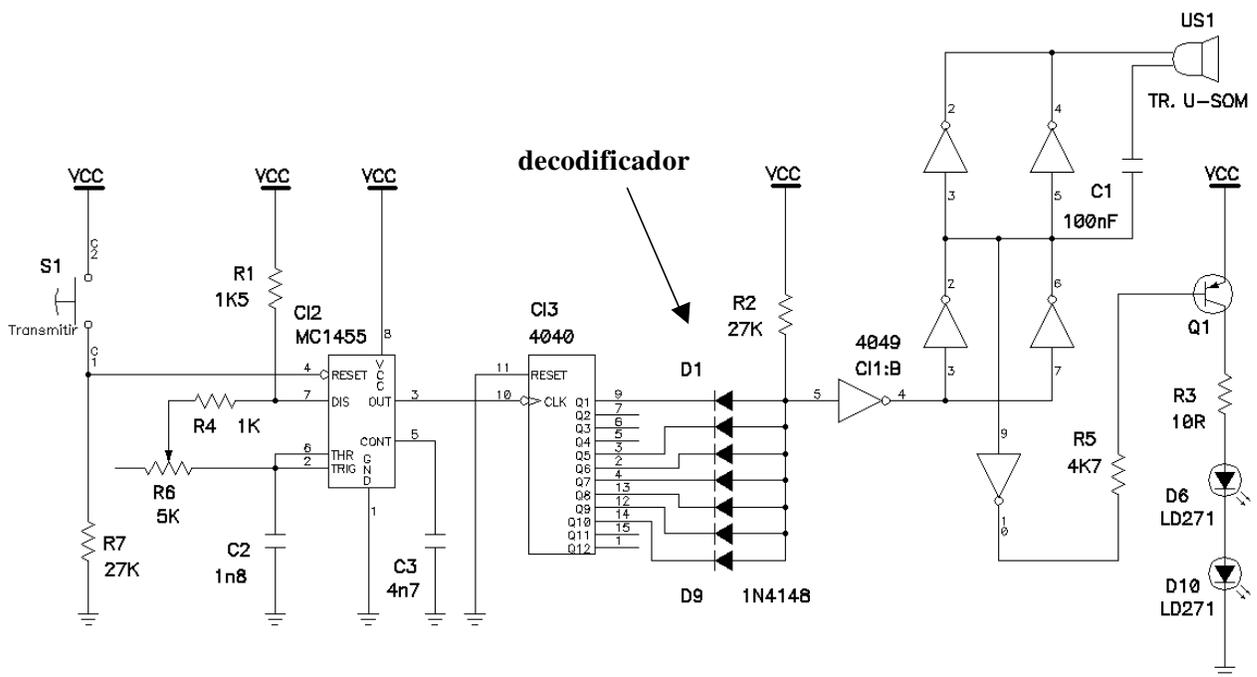


Figura 2.14 – Diagrama eletrônico do transmissor

No circuito do receptor, a borda de subida do 8º pulso (dos 8) do sinal de infravermelho é usada para marcar o instante de início de propagação da onda de ultra-som. Para isso é transformada em um breve pulso digital capaz de inicializar simultaneamente dos três temporizadores (contadores digitais) que fazem a contagem de tempo.

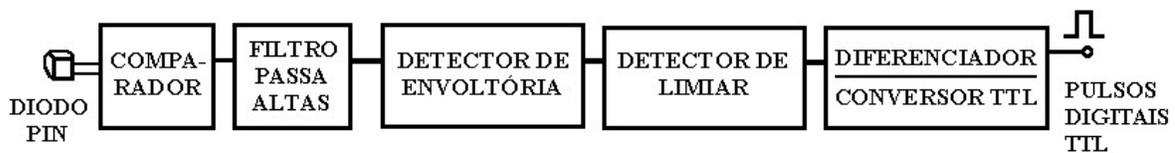


Figura 2.15 – Diagrama de blocos do circ. do sensor de infra-verm.

Este sinal é captado por um sensor de luz infra-vermelha, um diodo PIN, com máxima sensibilidade para luz em comprimento de onda na faixa de 950 nm. O dispositivo usado tem resposta casada com o led infra-vermelho do transmissor, e foi escolhido um tipo com ângulo de abertura de recepção bem amplo, para manter a flexibilidade de movimentos do usuário o mais próximo possível do que oferecem os sensores de ultra-som. O sinal do diodo PIN chega a um comparador com histerese, que converte em onda retangular os sinais recebidos em uma faixa ampla de intensidades. Um filtro passa-altas evita a captação de sinais muito abaixo dos 40 KHz (frequência de corte de 1,5 KHz), para evitar a interferência de iluminação artificial ou da luz do sol. Um seguidor de envoltória retira a componente tensão alternada de 40 KHz,

resultando em um pulso com duração igual ao agrupamento dos 8 pulsos do sinal original. A envoltória é comparada com uma pequena tensão de referência, detectando a borda de subida. Este sinal passa, finalmente, por um circuito diferenciador que limita a largura do pulso (positivo) que irá aos contadores, pois nível “1” segura os mesmos em estado de RESET (por um tempo menor que 500 nS) e deve voltar rapidamente ao nível “0” para permitir a contagem e não perder pulsos. Depois de ter suas tensões ajustadas aos níveis TTL, este sinal ganha o nome de “START”. Foi acrescentado um sinalizador de presença do sinal de saída com um led, para mostrar ao usuário quando o transmissor está bem orientado em direção ao receptor.

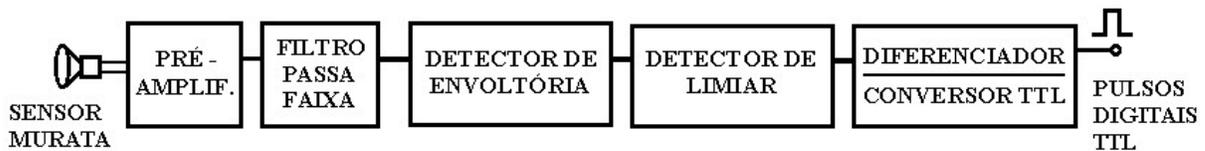


Figura 2.16 – Diagrama de blocos do circ. do sensor de ultra-som

O circuito dos receptores de ultra-som é bem parecido com o de infra-vermelho. Na entrada, o sensor é levado a um pré-amplificador com ganho de 200; passa em seguida por um filtro passa-faixa centrado em 40 KHz, por um seguidor de envoltória e chega a um detector com limiar variável, proporcional à distância do transmissor. O limiar de detecção da presença do sinal de envoltória tem um valor máximo no instante do início da transmissão (evitando detectar sinais de reflexão do trem de pulsos anterior atrasados) e diminui exponencialmente com o tempo ficando este estágio mais sensível se o sinal de ultra-som demora mais para chegar, por estar o transmissor mais longe e, conseqüentemente, o sinal mais fraco. A amplitude de sinal de ultra-som tem decaimento proporcional à exponencial da distância, de forma que este limiar proporcional ao tempo de propagação ajusta a sensibilidade automaticamente. O sinal de saída do detector passa, finalmente, por um circuito diferenciador que limita a largura do pulso (positivo) que irá aos contadores. Depois de ter suas tensões ajustadas aos níveis TTL, estes sinais ganham os nomes de “STOP1”, “STOP2” e “STOP3” (para 3 sensores de ultra-som). Foi acrescentado, para cada circuito de sensor, um sinalizador de presença do sinal de saída com um led, para mostrar ao usuário quando o transmissor está bem orientado em direção ao receptor.

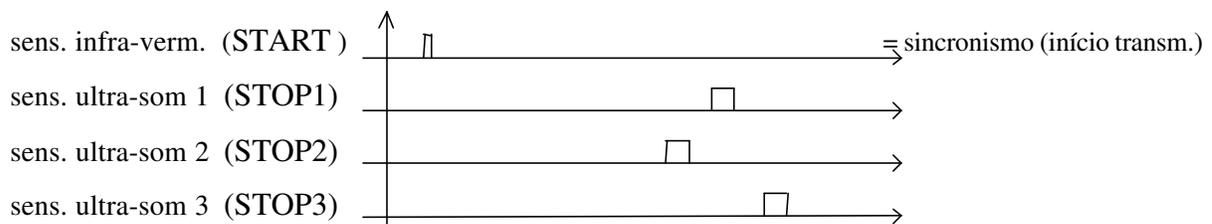


Figura 2.17 – Formas de ondas start e stop1 a 3

O módulo digital recebe os 4 sinais da figura 2.17 para proceder com a contagem de três intervalos de tempo usando contadores digitais, e posteriormente permitir sua leitura por meio de uma interface compatível com a porta paralela padrão Centronics (usada geralmente para comunicação entre computadores e impressoras). Estes valores digitais correspondentes aos tempos serão lidos por um computador, processados e organizados de uma forma conveniente dependente da aplicação.

Na prática, só existe um contador de 16 bits, e as contagens são capturadas em três registradores de 16 bits quando são detectados os sinais de ultra-som nos sensores correspondentes. O contador é reinicializado (levado ao módulo zero) com o sinal de START em nível “1”, sendo liberado para contar pulsos de clock quando este mesmo sinal voltar ao nível “0”. Os pulsos de clock referidos aqui derivam de um oscilador a cristal de 8 MHz, e, usando um contador binário para sua divisão por 2, 4, 8, ou 16, serão disponibilizadas as frequências de 8, 4, 2, 1, e 0,5 MHz para experimentações variando a resolução da medição de tempo.

Um a um vão sendo capturados os três tempos, e ao capturar o último, é gerado um sinal de aviso para que os registradores sejam lidos. Este sinal foi usado para gerar um pedido de interrupção reservado para a porta paralela, para que o computador pare o que estava fazendo e leia os registradores. O contador continua a totalizar o tempo até que receba nível “1” em START novamente. As relações entre este clock e o sinal de ultra-som determinam o módulo dos contadores, da seguinte forma:

$$t_c = \frac{1}{f_c} = \frac{1}{2 \times 10^6 \text{ Hz}} = 500 \times 10^{-9} \text{ S}$$

$$cont_i = \frac{tof}{t_c} = \frac{dof}{340 \text{ m/s} \times 500 \times 10^{-9}} = \frac{d_{\text{sen.sor}_i}}{1,7 \times 10^{-4} \text{ m}}$$

$$d_{\text{sensor}_i} = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$$

$$tof = \text{time_of_flight}$$

$$d_{\text{sen.sor}_i} = \text{distância do transmissor ao sensor } i$$

Estes valores digitais correspondentes aos tempos serão lidos por um computador, processados e organizados de uma forma conveniente. Nas aplicações em que se deseja obter a avaliação explícita das distâncias, o primeiro passo a ser feito é calcular a distância d a partir do tempo t de propagação da onda entre um ponto e o outro multiplicando-se este valor pela velocidade c de propagação da onda neste meio. Neste trabalho considerou-se constantes

todos os parâmetros que alteram essa velocidade, como a temperatura, a pressão, a composição.

$$d = c \cdot t \quad [\text{metros}]$$

$$\text{onde } c = 340 \text{ m/s}$$

Em outras aplicações, sendo necessário aumentar a precisão de avaliação das distâncias, deve se levar em conta a temperatura do meio de propagação da onda, usando:

$$c = 331,31 \times \sqrt{1 + \frac{T}{273,15}} \quad [\text{m/s}], \text{ onde } T = \text{temperatura } ^\circ\text{C}$$

aproximação considerada em [56],

ou

$$c = \sqrt{\frac{\gamma \times R \times T}{M}}, \quad \text{onde } \gamma = \text{const. adiabática dos gases};$$

R = const. universal dos gases perfeitos ;

M = massa molecular ;

T = temperatura $^\circ\text{C}$

de cunho bem teórico , ou

$$c = 331,5 + 0,61 \times T \quad [\text{m/s}], \text{ onde } T = \text{temperatura } (^\circ\text{C})$$

aproximação vista em [53]

O processamento que vem em seguida depende da aplicação e, neste trabalho, o próximo passo foi o cálculo das coordenadas x,y,z do transmissor.

2.5 Aquisição de Dados e Cálculo de Posição em Três Dimensões

Os sinais obtidos pelo chamado módulo analógico determinam o início e o fim dos tempos de propagação para os três sensores (possivelmente 4 sensores). A seguir, no módulo digital, estes tempos serão medidos pela contagem do número de pulsos de uma base de tempo gerada por um oscilador a cristal. Estes valores digitais correspondentes aos tempos

serão lidos por um computador, processados e organizados de uma forma conveniente dependente da aplicação.

Pelo que se vê na literatura [56,61], há um limite teórico para a resolução da medida de tempo de propagação de ondas sonoras, relacionado com a frequência de oscilação, em torno de $\lambda/2$. No caso dos transdutores usados neste trabalho, de 40 KHz, o limite teórico é :

$$\frac{\lambda}{2} = \frac{1}{2} \times \left(\frac{c}{f} \right) = \frac{1}{2} \times \left(\frac{340\text{m/s}}{40\text{KHz}} \right) = 4,25\text{mm}$$

onde :
 λ = comprimento de onda
c = velocidade do som no ar
f = frequência de oscilação

Uma onda quadrada com frequência de 2 MHz corresponde a ciclos com período de 500ns. Neste período, uma onda sonora percorre no ar $500 \times 10\text{E-}9[\text{s}] \times 340[\text{m/s}] = 0,17 \text{ mm}$, distância 25 vezes menor que a resolução teórica (4,25 mm), e que portanto não justificaria o uso de tal base de tempo em uma análise menos ponderada. Esta estimativa teórica não inclui alguma informação contida na região de incerteza da medida, que pode ser extraída dependendo da forma de transdução (com filtragem de altas frequências, muito comum, e usada neste trabalho) e de processamento dos dados de medição (extração de média, por exemplo, como feito neste trabalho). Levando estes procedimentos em consideração, não se altera o limite físico da medida, mas se aumenta a resolução da informação obtida no processo de medição. Na prática, se percebeu, um pequeno aumento da amplitude de oscilações na medida de posição do transmissor em um ponto fixo quando se aumenta a base de tempo de 500 ns para 1 μs , portanto foi escolhida a frequência que elimina esta possível fonte de imprecisões.

São contados então quantos pulsos de 500 ns ocorreram entre o início e o fim dos intervalos de propagação do sinal de ultra-som para os três sensores individualmente. Usando contadores de 16 bits, pode-se alcançar tempos de propagação de $65535 \times 500 \text{ ns} = 0,032768 \text{ s}$. Multiplicando este valor pela velocidade aproximada do som no ar, chegaremos ao valor do limite de distância que se pode medir nessas condições, de 11,141 metros, quase 4 vezes maior que o alcance obtido com o módulo analógico (com os transmissores e sensores de ultra-som e infra-vermelho usados). O contador com 16 bits tem portanto uma sobra de módulo de 2 bits, prevendo as experimentações com maiores resoluções de medida, usando transdutores de maior frequência.

O controle do sequenciamento para a aquisição de medidas é feito por meio de lógicas integradas em uma EPLD de 68 pinos. A integração de praticamente toda a parte puramente digital do circuito em um só circuito integrado permitiu a confecção de um dispositivo compacto e de baixo custo, características essenciais para torná-lo útil como ferramenta de desenvolvimento em outros trabalhos de pesquisa, como os sugeridos no fim deste texto. A disponibilidade de espaço para inclusão de lógica adicional na EPLD que comportou o

circuito permitiu algumas ampliações na forma de leitura dos contadores e de outros sinais internos, flexibilizando a interface com o computador a ser usado, pois permite usar “*handshaking*” baseado em interrupção e possibilita a ampliação da frequência de clock para experimentar com transdutores de maior ou menor frequência. Além disso, possibilitou uma grande simplificação do *lay-out* da placa de circuito impresso do módulo digital. A possibilidade de substituir boa parte do circuito digital pela simples reprogramação da EPLD aumentou a flexibilidade de confecção de um protótipo, podendo ser alterada toda a estratégia de controle e leitura dos dados, a sua precisão (em número de bits), e os terminais de diagnóstico para depuração do sistema. Veja a figura 2.18.

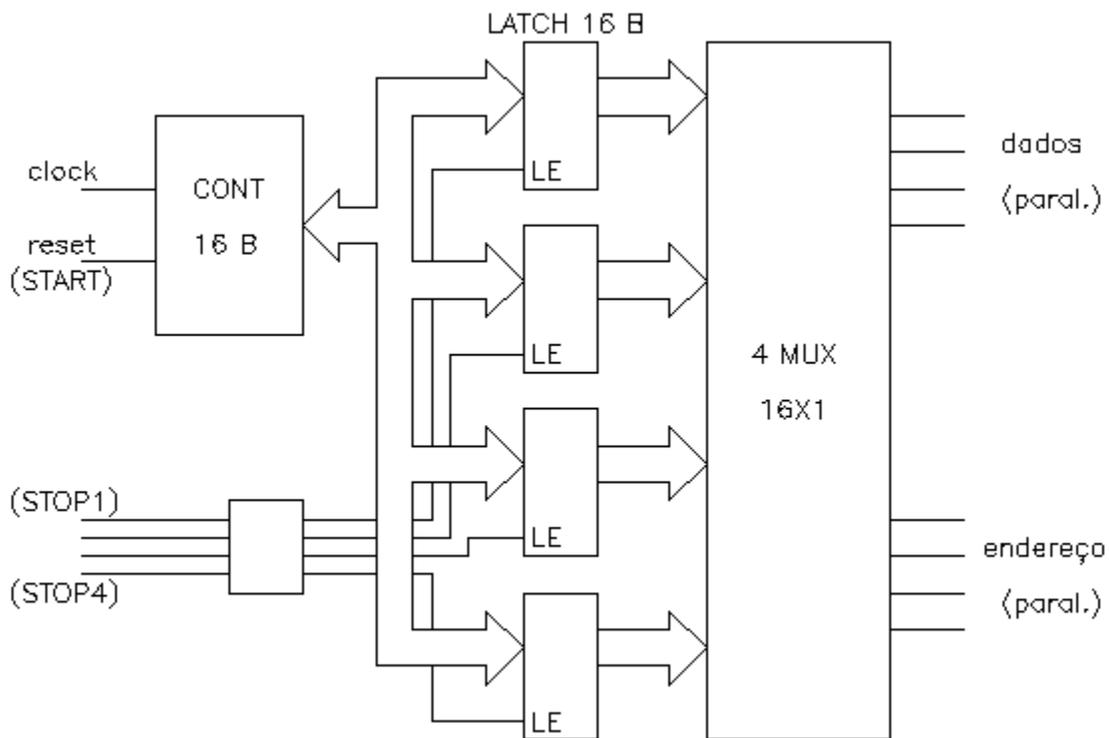


Figura 2.18 – EPLD: diagr. Interno em blocos

O circuito interno completo da EPLD inclui :

-] - um contador síncrono de 16 bits com entradas para carga de valor inicial (para uma possível compensação automática de erros de offset comuns aos 3 ou 4 sensores)
- 4 registradores de 16 bits para captura dos valores de totalização do contador (prevendo o uso do quarto sensor)
- biestáveis para controle da captura dos tempos de propagação

- lógica para geração de interrupção no momento em que se completa a captura de todos os 4 tempos
- lógica para endereçamento e controle de leitura dos registradores de captura em parcelas de 4 bits, em um total de 16 *nibbles*.
- foram disponibilizados alguns terminais de diagnóstico para depuração do sistema: os 16 bits de saída do contador de 16 bits, entregues internamente aos registradores de captura, os pulsos de captura dos 4 registradores, uma saída de estouro dos contadores.

A leitura destes *nibbles*, que correspondem aos valores dos tempos de propagação do ultra-som, é feita através da interface paralela padrão Centronics de um computador pessoal. Os modos mais rápidos disponíveis para transferência de dados por este meio chegam a taxas da ordem de megabytes por segundo (modos EPP, ECP). A fim de não correr o risco de limitar seu uso aos computadores mais modernos, e também pela simplicidade de programação e pela baixa velocidade de comunicação requerida, a opção foi feita pela leitura de 4 bits a cada acesso, através dos bits de status (endereço 379h). Os bits de dados de saída (endereço 378h) foram usados para selecionar qual dos 16 *nibbles* seria lido em seguida. O sinal de Acknowledge foi usado em sua opção de entrada para requisição da interrupção 7.

Um programa em linguagem C foi criado para fazer o controle do sequenciamento para a aquisição de medidas e da leitura dos dados, além de calcular as coordenadas dos pontos em 3 dimensões e pré-processar (com filtro de média) a sequência de pontos obtida (detalhes no próximo item deste capítulo). Foi compilado para execução no sistema operacional MS-DOS, por oferecer maior facilidade em curto prazo na implementação de serviço a interrupções. Neste programa foram integradas as seguintes funções:

- Organização da entrada de gestos pelo usuário

O programa pede ao usuário um conjunto de, no máximo, 6 caracteres, e uma faixa de valores numéricos com dois algarismos para comporem o nome de cada gesto com o fim de particularizar cada seqüência de gestos. Antes de cada gesto executado, o programa pede que o usuário pressione uma tecla, para então pressionar o botão de acionamento do transmissor e iniciar o gesto que não deve ultrapassar o tempo de aproximadamente 1 segundo e meio. O usuário é alertado se houver uma quantidade inaceitável de erros, provocados por exemplo, pelo mal direcionamento do transmissor, ou por soltar o interruptor do transmissor. Ele pode então parar o processo e recomeçar a seqüência a partir do gesto errado.

- Configuração da porta paralela para aceitar interrupções

Uma série de configurações são feitas no controlador de interrupções, como a sua prioridade, o endereço da rotina de serviço à interrupção, sua habilitação ou inibição em certos momentos estratégicos.

- Comunicação com o módulo digital do circuito dos sensores

Em resposta a uma interrupção que identifica a disponibilidade de um ponto adquirido, são gerados seqüencialmente endereços e sinais de controle para leitura de cada um dos 16 *nibbles* que compõem o conjunto de valores de tempo para o ponto e são convertidos os bits lidos em seus níveis lógicos originais (pois alguns bits são invertidos pela interface paralela).

- Sequenciamento de pontos dos gestos em matrizes

Cada nibble é guardado em uma matriz temporária onde cada linha corresponde a um ponto do gesto, cada linha com 16 *nibbles*, armazenados para posterior conversão. Este procedimento foi feito com atenção especial para a velocidade de execução, para diminuir os riscos de se comprometer a aquisição de dados em computadores mais lentos, já antecipando uma opção futura de controle deste sistema por pequenos microcontroladores de 8 bits.

Obs.: Após a aquisição de todos os *nibbles* correspondentes à seqüência dos pontos todos de um gesto, a matriz destes *nibbles* será varrida na mesma seqüência de escrita e são feitos os cálculos para obtenção, no final, de outra matriz com as coordenadas x,y,z.

- Cálculo das distâncias do transmissor até cada um dos sensores

Inicialmente os *nibbles* de um ponto são separados em palavras de 16 bits e convertidos de binário para decimal. A partir destes módulos dos contadores já em base decimal, multiplicando pela base de tempo usada de $1/(2 \text{ MHz})$, são obtidos os tempos de propagação do ultra-som. Cada valor é multiplicado então pelo valor aproximado da velocidade de propagação do som no ar, de 340 m/s para cálculo das distâncias do transmissor até cada um dos sensores em décimos de milímetros.

- Filtragem da seqüência de pontos

Para cada ponto é calculada a média com o ponto anterior, para obter uma curva mais suavizada no final. O objetivo inicial deste procedimento era de se evitar a necessidade de aumentar as distâncias entre os sensores, como descrito no item 2.3, e foi mantido mesmo depois da decisão pelo aumento das distâncias, por ser um procedimento simples e de grande efeito positivo.

- Cálculo das coordenadas x,y,z a partir das distâncias obtidas

Para cada medida obtida de três distâncias são calculadas as coordenadas x,y,z do ponto, obtendo-se a interseção entre três esferas de centros nos sensores e que têm em comum um ponto de sua superfície, nas coordenadas do transmissor, como descrito no item 2.3.

- Gravação, em disco, de arquivos com nomes seqüenciais.

Assim que o programa calcula o trio de coordenadas de mais um ponto, esta informação poderia ser usada de imediato para posicionamento de um cursor na tela de um computador, e/ou fazer com que um sistema tenha reações de acordo com o movimento do transmissor. Ao término de um gesto completo poderia se tentar seu reconhecimento e como consequência partir para a execução de uma tarefa associada a este gesto (1 gesto = 1 comando). É desta forma que se imaginou a utilidade prática deste projeto, com a ajuda de *softwares* interativos, para explorar as possibilidades da interação com a máquina por meio de gestos. Para a aplicação tomada como exemplo neste trabalho, o *software* ora descrito termina por gravar uma linha a mais em um arquivo, que no final terá armazenado um gesto completo em uma matriz de 128 linhas de 3 colunas. O nome deste arquivo pode identificar o usuário, o tipo de gesto que está sendo executado, e um número seqüencial de dois dígitos gerado automaticamente, como por exemplo: gj32.dat => usuário g (de Gilberto) , gesto j (letra J) , número seqüencial 32.

Para testes de reconhecimento de gestos por redes neurais, foram gerados 50 gestos (XX01.dat até XX50.dat) de cada um de 8 tipos, repetidos por dois usuários, num total de 800 gestos, ou seja, 800 arquivos .dat. De posse destes arquivos, posteriormente foram criados arquivos que agrupavam todos os gestos de cada um dos dois usuários em arquivos separados, e também do conjunto de todos os gestos dos dois usuários em um terceiro arquivo, para facilitar sua manipulação pelos programas desenvolvidos em Matlab para simulação das redes neurais.

A velocidade de propagação do som no ar foi considerada pelo valor aproximado de 340 m/s. Havia sido inicialmente prevista a necessidade de uma medida absoluta mais exata, considerando que a velocidade de propagação do som no ar varia com a temperatura do meio em que se propaga. No cálculo deste valor mais exato, pode-se fazer alguma simplificação, como se vê em alguns trabalhos publicados

Para tanto foi incluído no módulo digital do circuito dos sensores um medidor de temperatura com escala de 0 a 70 graus *Celsius*, lido também por meio de um bit da porta paralela, multiplexando seus 8 bits. A normalização das medidas dos gestos (para seu reconhecimento por redes neurais) fez com que se tornasse importante a precisão das proporções entre as distâncias medidas (e conseqüentemente as calculadas), sem que houvesse necessidade de precisão na medida absoluta de cada ponto, ficando sem sentido a medida de temperatura ambiente. Esta medida alteraria um pouco um fator de escala posteriormente padronizada para uma unidade, com a normalização. Mesmo assim este circuito foi mantido no protótipo para prever outras aplicações sem normalização.

Nas próximas duas páginas (figura 2.19) é mostrado o fluxograma do programa desenvolvido para aquisição de gestos, para maior clareza. No Apêndice A se encontra a listagem completa do mesmo programa, desenvolvido em linguagem C.

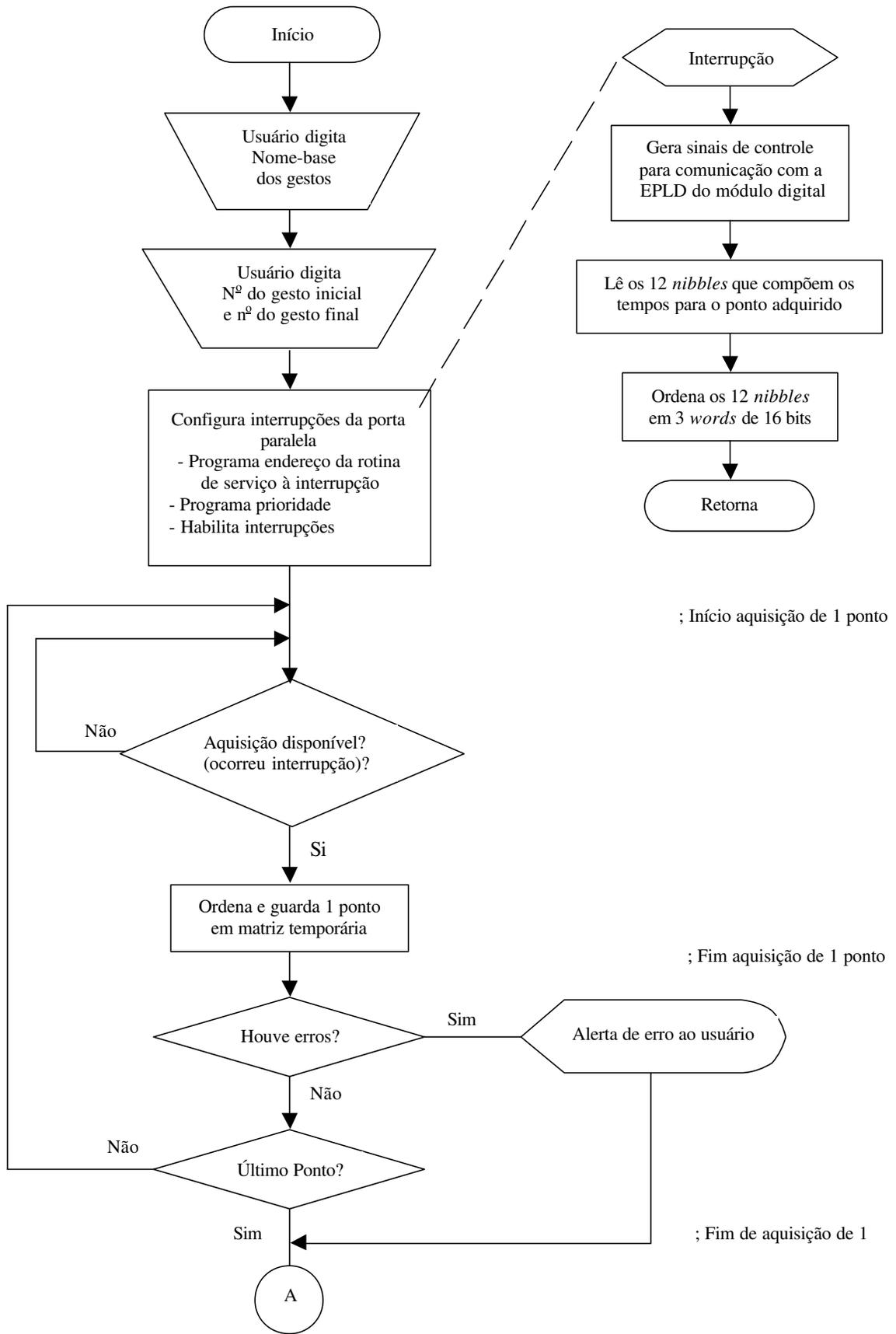


Figura 2.19 – Fluxograma do programa de aquisição de gestos

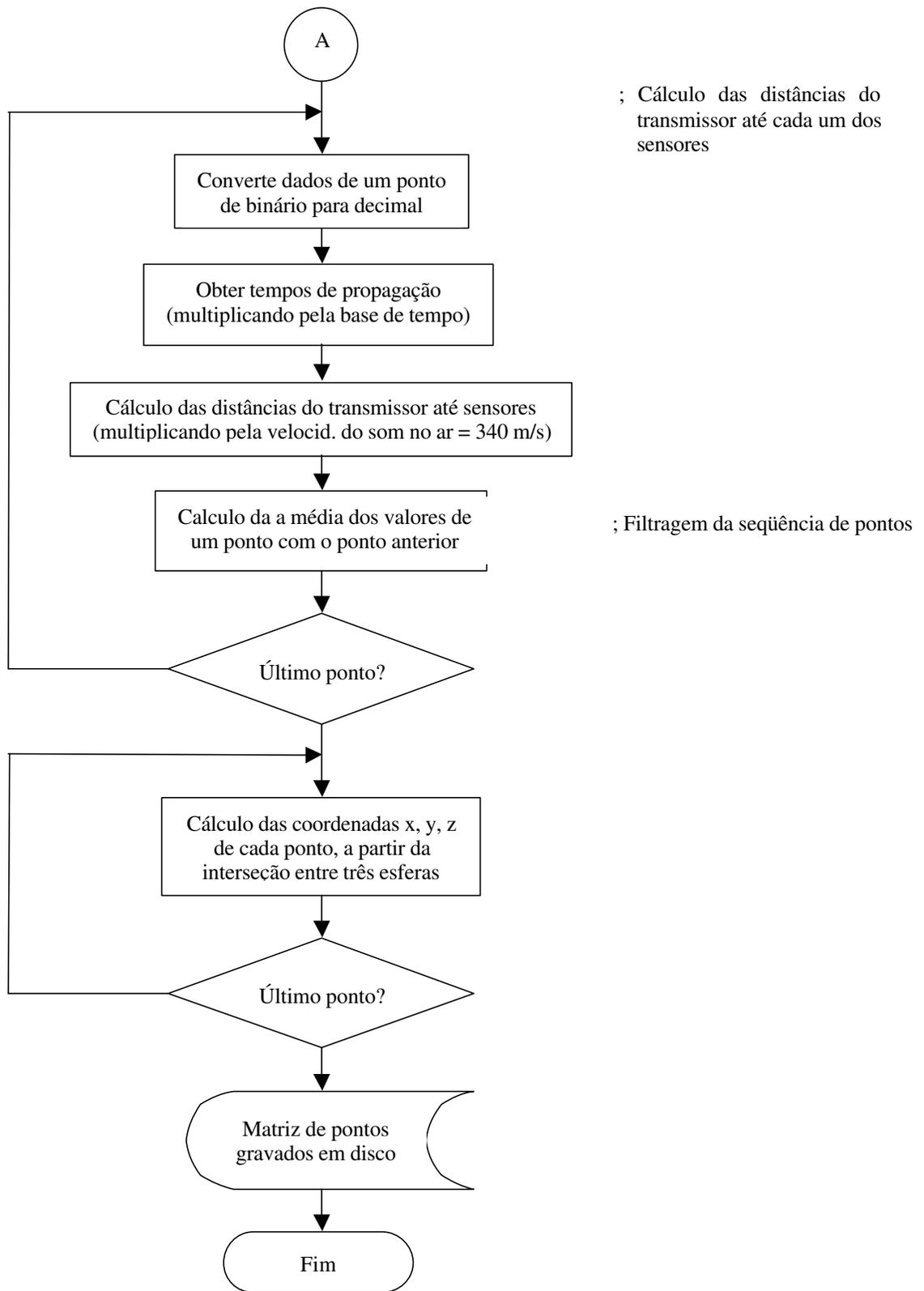


Figura 2.19 (cont.) – Fluxograma do programa de aquisição de gestos

Capítulo 3 - Identificação de Gestos

3.1 Delimitação de Gestos

O sistema eletrônico desenvolvido e o *software* de aquisição/armazenamento de seqüências de pontos descritos no capítulo 2 tiveram como objetivo criar uma plataforma para experimentação da interação do Homem com uma máquina através de gestos. Este trabalho considera, para condução de alguns testes de aquisição e reconhecimento o uso de gestos executados com restrição de tempo em 1,64 segundos. Isto foi feito para adiantar a obtenção de resultados, deixando o desenvolvimento de *softwares* mais complexos de controle de aquisição de gestos para aplicações futuras, que permitam outras formas de uso com restrições menos rígidas, ou mais confortáveis para o usuário, ou até sem restrições de tempo.

Para a aplicação tomada como exemplo neste trabalho, grupos de gestos são amostrados em seqüência para posterior uso no treinamento e avaliação de redes neurais. Um gesto tem seu início no instante em que é recebida a primeira amostra de um ponto, e termina quando chega a 128ª amostra de pontos. Cada ponto do gesto, correspondente a uma linha com 12 *nibbles*, é armazenada em uma matriz para posterior conversão. Após a aquisição de todos os pontos de um gesto, a matriz será varrida na mesma seqüência de escrita, reagrupando os *nibbles* em 3 palavras de 16 bits, convertendo-os em binário, filtrando-os com a extração de média e fazendo os cálculos geométricos para obtenção, no final, de outra matriz de 128 linhas de 3 colunas com as coordenadas x, y, z dos pontos (veja fluxograma nas páginas anteriores). Esta é então gravada em um arquivo, cujo nome identifica o usuário, o tipo de gesto que está sendo executado, e um número seqüencial de dois dígitos gerado automaticamente.

Antes de serem usados para treinamento e reconhecimento por uma rede neural, os gestos armazenados nos arquivos serão pré-processados : sofrem um processo de reamostragem, passando de um número qualquer de pontos do gesto original (128) para um número conveniente de pontos (32, por exemplo), igualmente espaçados entre si, o que permite implementar no futuro a delimitação de gestos com número variável de pontos amostrados (tempo de gesto variável).

3.2 Redes Neurais e Classificação de Seqüências

3.2.1 Redes Neurais Artificiais

A construção de redes neurais artificiais para processamento de dados e sinais foi inicialmente motivada pelo desafio de estudar e imitar o comportamento do cérebro, que

contém um sistema de processamento de dados que funciona de uma maneira completamente diferente dos computadores digitais convencionais. O córtex cerebral é constituído de uma quantidade enorme de unidades de processamento de pequena capacidade (neurônios) , interligados em uma rede complexa cuja estrutura contém, nas ligações (sinapses), memórias de parâmetros estatísticos (pesos sinápticos) que encerram um conhecimento adquirido pela repetição de exemplos (e não por memorização dos casos individuais ou por programação). Esta característica destaca sua utilidade em aplicações onde há pouco ou nenhum conhecimento a respeito de um problema a ser resolvido, mas existem disponíveis conjuntos de dados com pares de problema - solução (ou entrada - saída desejada) típicos do sistema em questão.

A forma de processamento é mais parecida com processamento de sinais (filtros adaptativos) do que com processamento de símbolos e números. Seu bom desempenho vem do paralelismo massivo com que são realizadas as tarefas e de sua capacidade de aproximar não-linearidades . Outras características estratégicas das redes neurais são :

1- a capacidade de lidar (processando, após treinamento) com dados de entrada nunca vistos e que são considerados (e produzem resultados coerentes) pela generalização a partir dos dados de treinamento ;

2- sua robustez na presença de falhas de *hardware*, pois a mudança (ou anulação) de um neurônio em meio a milhões tem pouco significado, afetando os resultados para alguns poucos dados de entrada ;

3- o ajuste de parâmetros (treinamento) é uma tarefa relativamente simples em comparação às ferramentas de processamento similares, podendo ser mais facilmente executada sem a interferência de especialistas em programação, processamento de sinais, ou inteligência artificial, pois imitam as características estatísticas desejadas sem a necessidade de conhecimento explícito de parâmetros estatísticos, como a distribuição de probabilidades.

Vêm sendo construídos, em centros de pesquisa de todo o mundo, simuladores de redes neurais, máquinas com o objetivo de operar de uma forma similar à do cérebro na execução de tarefas específicas, usando componentes eletrônicos dedicados ou *software* em computadores digitais. Os modelos matemáticos de neurônios usados nas simulações são muito limitados em relação ao que já se conhece da funcionalidade dos neurônios naturais, mas permitem avaliar vários aspectos de seu funcionamento, os métodos de aprendizagem e as possíveis topologias de redes. Como ferramentas, os modelos de neurônios e de redes neurais oferecem algumas vantagens substituindo os modelos tradicionais de processamento estatístico de sinais, por lidarem bem com processos não lineares, não estacionários e até não determinísticos (caóticos) [29]. Além disso, é desafiante entender o funcionamento de uma máquina muito mais eficiente que os computadores digitais: o cérebro gasta 10^{-16} joules por operação contra 10^{-9} nos melhores computadores conhecidos (gasta dez milhões de vezes menos energia para tarefas similares) [26,28].

As redes neurais têm sido objeto de estudo multidisciplinar, envolvendo pesquisadores das áreas de informática, engenharia eletrônica, medicina, biologia, física, filosofia, havendo

uma complementaridade entre estes estudos, pois os avanços obtidos em uma área sugerem novos caminhos para as outras, no desenvolvimento de modelos mais fiéis ou mais úteis dos neurônios, das topologias de redes e dos métodos de aprendizado. É o caso do estudo do aparelho visual de animais sugerindo a inclusão de etapas de pré-processamento de imagens feito por tipos especializados de neurônios. Outras áreas de pesquisa (economia, música, geologia, telecomunicações, etc.) fazem uso das capacidades de processamento das redes neurais na solução de problemas específicos, remetendo a novas formas de uso, isoladamente ou como parte de um processo maior. Um bom exemplo de colaboração no desenvolvimento vem do estudo de técnicas de processamento de sinais: o algoritmo de atualização de parâmetros de filtros adaptativos pelo método dos mínimos quadrados (LMS) tem sido usado com sucesso no aprendizado das redes neurais ; em contrapartida, a adequação das redes neurais ao emprego de processamento paralelo tem sido uma boa opção para aplicações de alto desempenho no processamento de sinais.

A “programação” destes “processadores neurais”, estruturas de processamento paralelo altamente distribuído, é feita modificando os pesos das conexões entre processadores de forma ordenada, reforçados ou diminuídos objetivando a aproximação gradativa de uma solução aceitável (não exata). Este processo é chamado de aprendizado ou treinamento, e a solução aceitável mencionada que se obtém no final pode ser vista como uma função de transferência. Esta relação entrada-saída pode ser usada de várias maneiras, por exemplo como um filtro adaptativo, que tem seus parâmetros constantemente alterados de acordo com a evolução das variáveis de um processo e seguindo algum critério. Desta forma a rede aproxima a função de transferência do filtro desejado, de forma supervisionada ou não. Pode ser usada também para fazer a ativação (atribuir o maior valor possível) uma de “n” saídas (classes) quando um conjunto de dados é apresentado às entradas. Esta tarefa é denominada classificação, ou reconhecimento de padrões (implementado neste trabalho), pois determina a probabilidade de uma amostra pertencer a uma das possíveis classes ou padrões conhecidos (aprendidos, ou treinados). A particularidade do treinamento para classificação está na configuração das saídas, pois para cada configuração de entrada específica uma delas tenderá a assumir o valor máximo, enquanto as outras saídas tenderão para o mínimo.

A análise da conveniência de cada processo de treinamento para um determinado problema, e o formato das informações disponíveis sobre o sistema, determinam a escolha da arquitetura de uma rede neural a ser usada. Quanto à forma de treinamento, podemos separar as mesmas em dois grandes grupos:

1- Redes com aprendizado supervisionado: A arquitetura deste tipo de redes permite a apresentação de exemplos de dados de entradas associados com as saídas desejadas correspondentes. Para cada amostra de dados nas entradas, a rede faz associações entre as mesmas usando os parâmetros correntes (pesos sinápticos) avaliando os valores das saídas. As diferenças entre os valores de saída obtidos e os desejados são usados para calcular o quanto será modificado em cada peso sináptico (inicializados de forma aleatória) de cada neurônio, em direção a minimizar estas diferenças. É como se houvesse um professor, orientando o aprendizado passo a passo, caracterizando o aprendizado supervisionado, também conhecido aprendizado associativo.

2- Redes com aprendizado não supervisionado: Quando não há exemplos disponíveis com pares de correspondências entrada-saída, mas existem amostras dos dados que se deseja processar, o aprendizado será não-supervisionado. Consiste em levar os dados de exemplo sucessivamente às entradas, de forma aleatória, e progressivamente capturar as regularidades estatísticas destes dados, agrupando-os de acordo com suas similaridades, obedecendo a algum parâmetro comum (critério de mapeamento), sem que no final se tenha uma atribuição de nomes ou valores que possam identificar cada um dos grupos de dados. Muitas das soluções de engenharia têm exigido adaptação não supervisionada, o que tem sua utilidade na classe de problemas de identificação de seqüências sem o conhecimento de uma medida de referência, obtendo apenas informações relativas, ou seja, descrevendo matematicamente o percurso, sua dinâmica, sem medidas quantitativas absolutas, avaliando e corrigindo erros por critérios relativos.

Quanto à forma de conexão entre os neurônios, as redes neurais podem ser classificadas como recorrentes (com realimentação) ou não recorrentes. A figura 3.1 mostra uma rede com realimentação, que se assemelha muito com os filtros digitais realimentados, pois algumas das entradas são usadas para amostrar o estado anterior das saídas [28].

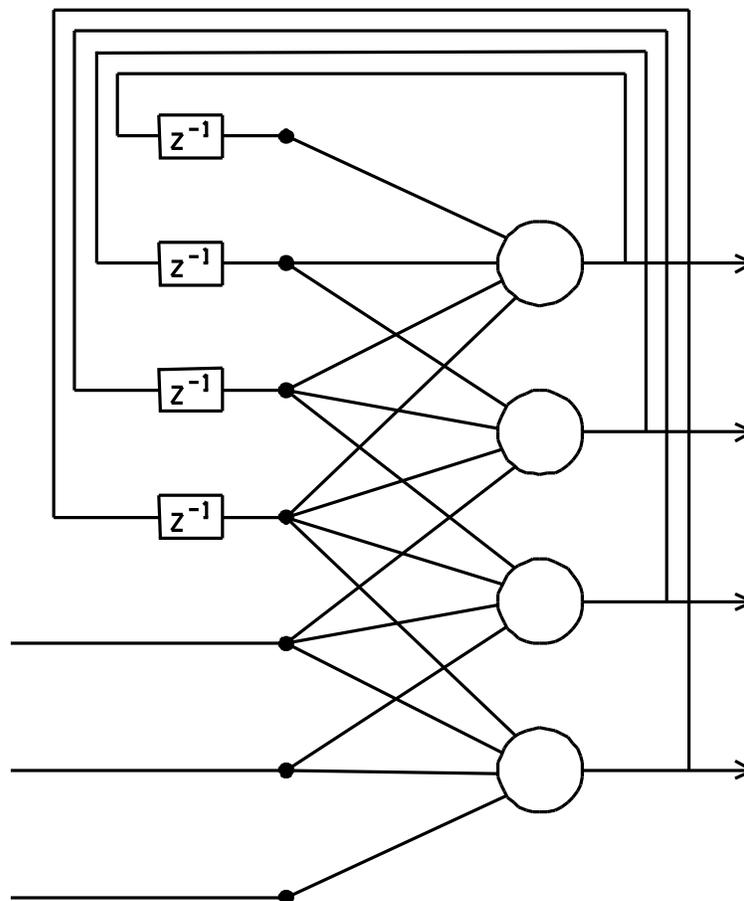


Figura 3.1 – Rede neural recorrente

Podemos classificar as redes neurais também pela sua arquitetura: temos as reticuladas (organizadas em matrizes de neurônios, com todas as entradas de cada neurônio ligadas a todas as entradas da rede, como na fig. 3.2) ou estruturada em camadas, onde só a 1ª camada tem acesso às entradas (fig. 3.3).

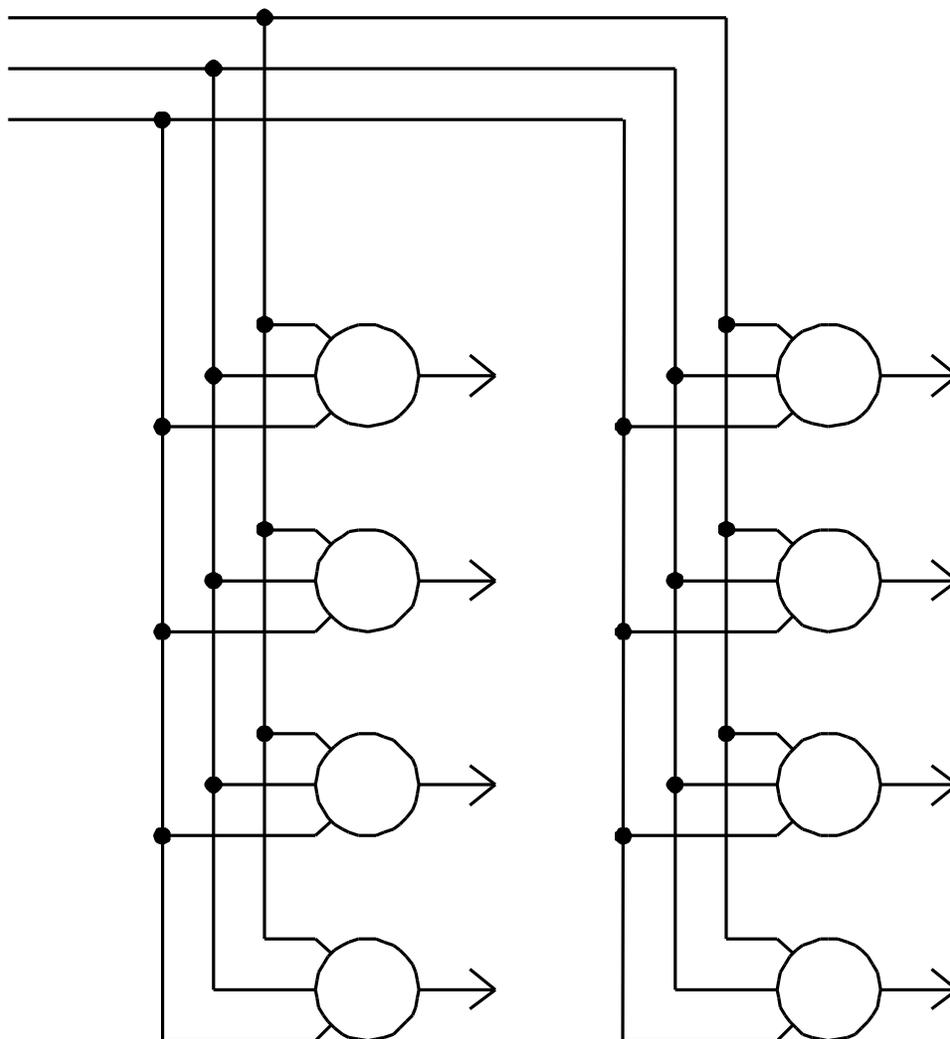


Figura 3.2 – Rede neural reticulada

Estas últimas podem ser totalmente ou parcialmente conectadas entre as camadas para economia de custo computacional na extração de características reconhecidamente localizadas das amostras de entrada (figura 3.3).

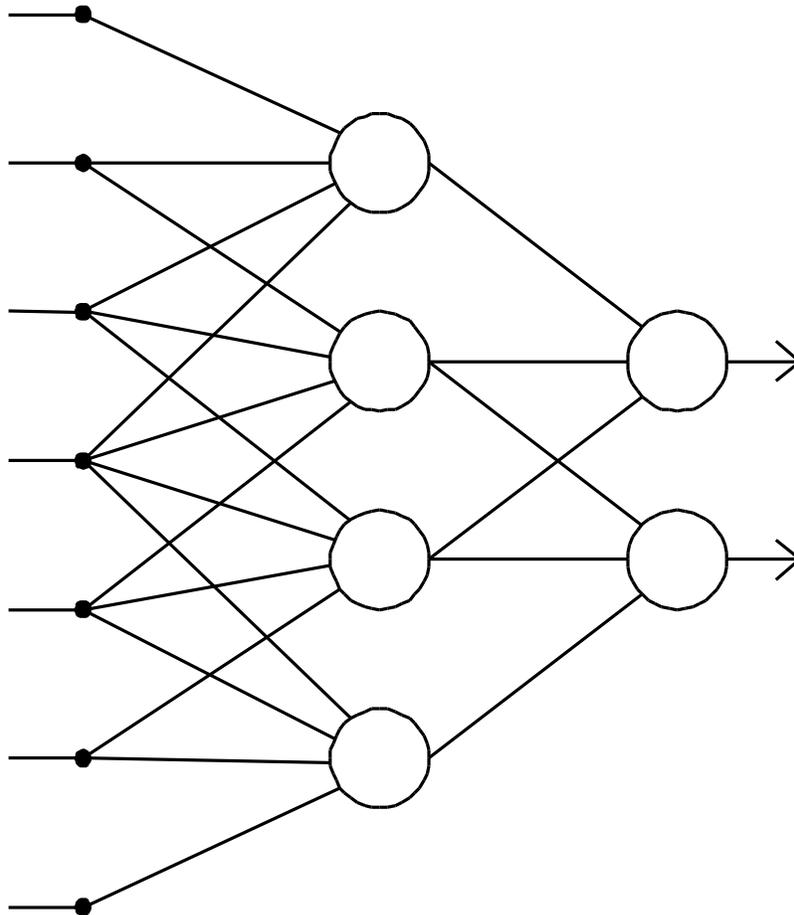


Figura 3.3 – Rede neural em camadas, parcialmente conectada, não recorrente

Neste trabalho foram implementadas redes neurais do tipo Perceptron Multicamadas, totalmente conectadas, como a da figura 3.4, com treinamento supervisionado usando o algoritmo de *backpropagation* [28,34] para classificação de gestos (detalhes, mais adiante). São constituídas de uma camada de entrada (pontos pretos na figura), camada de saída e camadas escondidas (das quais fazem parte todos os neurônios exceto os de saída).

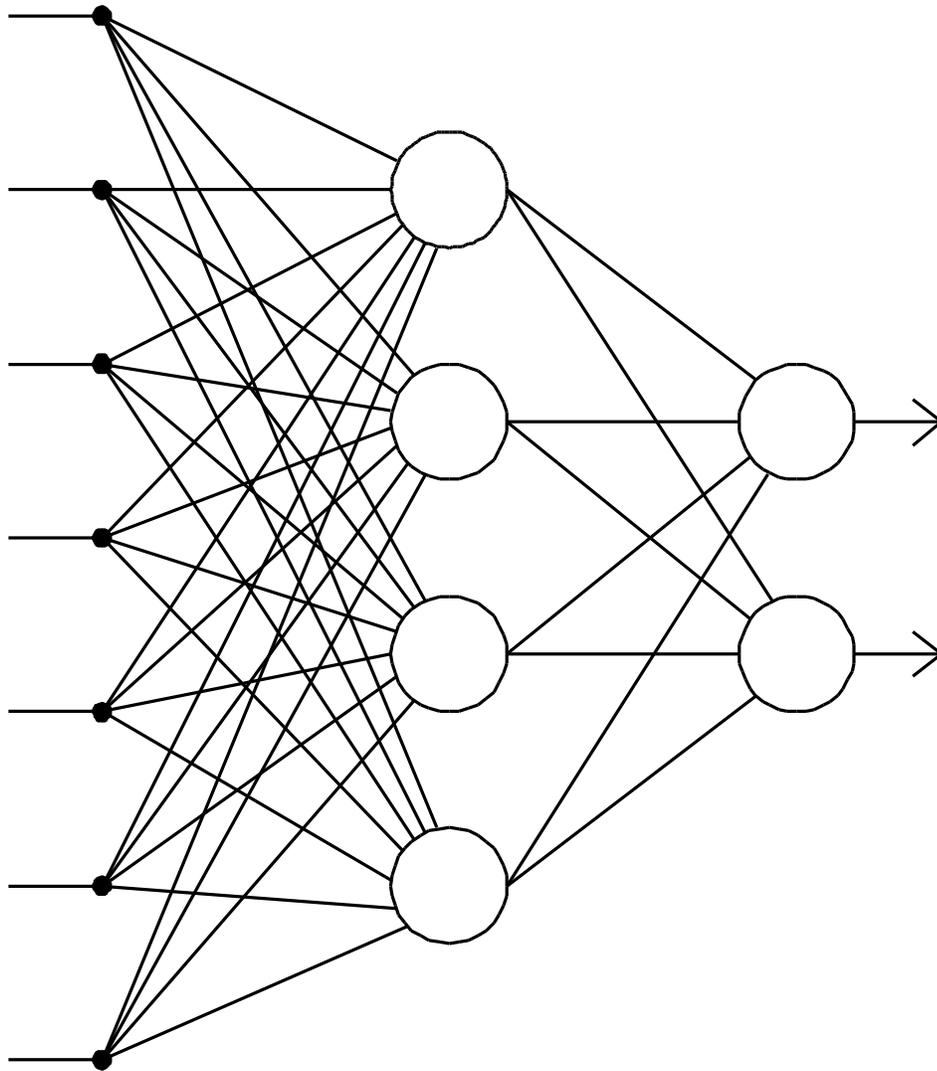


Figura 3.4 – Rede neural em camadas, totalmente conectada, não recorrente

Nem sempre a melhor solução para um problema de processamento de dados é a que usa o modelo de neurônio ou da topologia de redes neurais que mais se aproxima do modelo do cérebro. Por isto, existem vários tipos de redes neurais artificiais, cada um adequado a uma classe de problemas, por conseguir extrair os elementos mais relevantes do processo em que se insere.

3.2.2 Redes Neurais na Aproximação de Funções

Se o problema a ser resolvido por uma rede neural é a aproximação de funções, já se tem a restrição pelo uso de treinamento supervisionado, que parte da apresentação de exemplos de dados de entradas associados com as saídas desejadas correspondentes. Para cada amostra nas entradas, a rede avalia os valores das saídas e usa as diferenças entre os valores de saída obtidos e os desejados para calcular o quanto será modificado em cada peso sináptico de cada neurônio, em direção a minimizar estas diferenças, objetivando a aproximação gradativa da relação entrada-saída da função, que pode ser não linear. Um tratamento mais formal da capacidade das redes neurais na aproximação de funções pode ser visto em [28]. Do processo de treinamento se obtém no final, por exemplo, a função de transferência de um filtro, ou o mapeamento das coordenadas retangulares de um ponto a partir das distâncias até três pontos conhecidos (implementado pelo autor com redes neurais em um trabalho não publicado, semelhante ao visto em [14], ou por cálculos geométricos, visto neste trabalho).

Os tipos de redes neurais mais usados na aproximação de funções são as conhecidas como Perceptron Multicamadas e RBF (*Radial Basis Function*), muito semelhantes em sua estrutura. Ambas são redes não lineares dispostas em camadas, mas com diferenças significativas de funcionalidade, principalmente porque na RBF cada neurônio corresponde a uma função que constitui, junto aos outros da única camada escondida, uma base de funções com dimensão tão grande quanto necessária para aproximar a função de interesse [28,34].

As redes usadas na implementação de funções de transferência de filtros devem ter, como no caso destes, a capacidade de captação e modificação das relações temporais entre as entradas. Isto poderia ser implementado diretamente em arquiteturas de redes neurais dinâmicas, tendo atrasos unitários de amostragem embutidos entre as camadas da rede [28]. Também os filtros espaciais encontram boas soluções nas redes neurais, como na implementação de algoritmos para *beamforming* (veja cap 2).

O sistema proposto pode ser usado na classificação de objetos em movimento, usando padrões deste movimento para diferenciar um objeto de outro. Por outro lado, os gestos de interesse deste trabalho poderiam ser vistos como uma seqüência de pontos amostrados em intervalos uniformes, de forma que uma rede neural deste tipo (com três entradas, uma para cada sensor) pudesse caracterizá-los pela dinâmica dos valores de tempos de propagação (aproximando suas funções temporais), que refletem a dinâmica dos movimentos da mão, dispensando a conversão de tempos de propagação para distâncias e destas para as coordenadas retangulares para a posterior classificação.

Como um valioso acréscimo, o processamento das amostras poderia incluir a filtragem espacial, para atenuar os ruídos e as interferências causadas por reflexões indesejadas, permitindo o aumento da sensibilidade e da precisão das medidas. Tudo isto aumentaria o custo do sistema, o que vai contra os propósitos deste trabalho. Além disto, os algoritmos de treinamento para estas redes, como o “*backpropagation through time*”, têm o inconveniente de demandarem muito tempo em seu treinamento, e a configuração proposta neste trabalho substitui tempo gasto para processamento e transferência de dados entre camadas da rede por

processamento em *hardware*, aumentando a eficiência que se pode alcançar com um sistema como este. Pode-se ver algumas aplicações muito interessantes de redes neurais dinâmicas de arquiteturas bem especializadas, como o modelo pulsado de neurônios para processamento na localização de sinais sonoros [39], e as redes criadas para reconhecimento de voz e interlocutor, além dos preditores adaptativos [29,30].

Por meio de filtros espaciais ("*beamforming*"), isoladamente, é determinada a direção de chegada de uma onda, e não se consegue diretamente a distância absoluta do objeto, obtida através da conjugação com outras técnicas de processamento de sinais. Várias destas técnicas se tornam apropriadas em aplicações onde se procura obter maior sensibilidade a algum aspecto do objeto e maior rejeição das interferências do ambiente [15]. Nos sistemas de SONAR para investigação no mar, por exemplo, redes neurais de alto desempenho têm sido usadas para predição do movimento das ondas, que tem grande parcela no modelamento matemático de distorções de sinais naquele ambiente. O fato das ondas do mar se caracterizarem por movimentos aproximáveis por funções caóticas determinísticas torna bem adequado o uso de preditores com redes neurais para compensação de distorções [34].

3.2.3 Redes Neurais na Classificação de Sequências

Quase todos os tipos de redes neurais podem ser usadas na classificação de conjuntos de dados, mas os que são treinados de forma supervisionada permitem a associação com classes bem determinadas, tornando-os adequados aos objetivos deste trabalho.

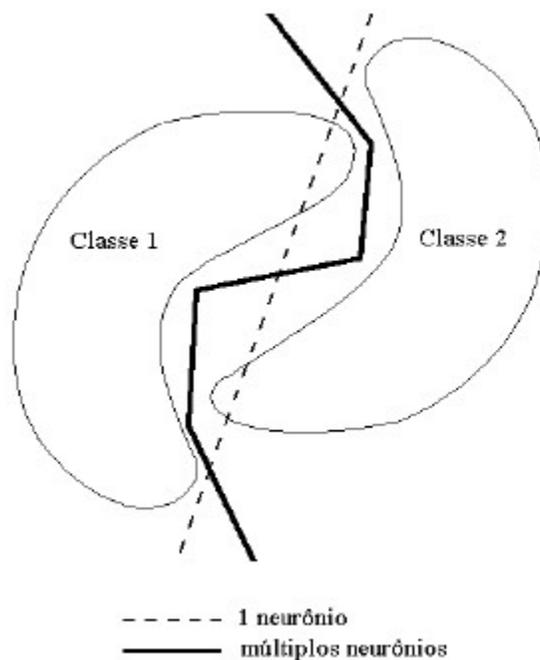


Figura 3.5 – Superfície de separação de classes

Um único neurônio “aprende” uma função linear de decisão quando é apresentado a dois conjuntos de dados de treinamento que sejam linearmente separáveis. Quando a separação não é perfeitamente linear, é encontrada a melhor solução, tendo em vista a limitação imposta. Para lidar com a separação de conjuntos de dados em três ou mais classes, emprega-se a conjugação de vários neurônios dispostos em camadas interligadas, permitindo a aproximação de superfícies de separação de classes bastante complexas (veja figura 3.5).

No contexto das redes neurais para classificação de seqüências, as redes do tipo Perceptron Multicamadas treinadas (de forma supervisionada) por um algoritmo de *backpropagation* se destacam por sua simplicidade de implementação, bom suporte teórico e prático pela disponibilidade de grande número livros, periódicos, exemplos com programas-fonte e tutoriais (incluindo material publicado na Internet) e pela vasta gama de implementações conhecidas com resultados muito positivos. Neste trabalho o interesse se volta então para as redes deste tipo, que permitem boas soluções na classificação de seqüências.

O problema apresentado à rede usada consiste na apresentação de seqüências de medidas de ângulos de deslocamento entre pontos amostrados que caracterizam gestos completos, para classificação dos mesmos. As questões de condicionamento numérico das medidas e as otimizações possíveis são analisadas no contexto do problema.

Os dados de entrada da rede usada correspondem a gestos dos quais se quer reconhecer o formato. Dois gestos de formatos idênticos podem ser executados com dinâmicas diferentes, um inicialmente rápido terminando em menor velocidade, outro começando lento e terminando rápido, ou mesmo com outras variações de velocidade durante o percurso. Como consequência, haverá variações no espaçamento entre amostras, levando a representações diferentes para gestos com formatos idênticos. Para evitar este problema, os gestos passam por uma reamostragem descrita no próximo item deste capítulo, para uniformizar os espaços entre amostras, e já providenciando outras tarefas de pré-processamento para adequar o formato dos dados à entrada da rede neural.

3.3 Pré-processamento e apresentação de dados à Rede Neural

Um procedimento de pré-processamento tem a importante função de facilitar, ou mesmo tornar possível, o procedimento consecutivo de processamento. Deve portanto garantir que componentes indesejados sejam retirados do sinal a ser processado, que a escala de valores esteja dentro da faixa de trabalho do processador, e se possível formatar os dados para aumentar a velocidade e/ou a precisão do processamento, sem que o conteúdo de informações do sinal seja corrompido [15].

Vários trabalhos que usam redes neurais para classificação empregam formas de pré-processamento específicas para o problema em questão, como no caso do processo de reconhecimento de caracteres manuscritos implementados em [20], conhecido como *Neocognitron*. Ele é composto de camadas de entrada para pré-processamento, com neurônios

treinados sem supervisão e camadas internas treinadas com supervisão. A arquitetura deste exemplo de rede neural reflete o conhecimento antecipado do processo, com camadas subsequentes parcialmente conectadas, que favorece a atenção às características localizadas, vistas a cada camada em subconjuntos mais amplos. O objetivo é a extração, não sensível à translação e à escala, de características locais dos caracteres, para posterior classificação do carácter. Nos trabalhos em que se usa esta primeira camada diferenciada, ela tem sido chamado de *feature extraction*.

A evolução do treinamento da rede pode obedecer algum critério, com o objetivo de favorecer a evolução em uma direção específica, também refletindo algum conhecimento antecipado do processo.

Outro exemplo de conjugação de arquiteturas de redes neurais vem da adaptação do sistema sensorial dos morcegos para identificação de alvos adequados, na caça noturna. Seu sistema de localização por ondas acústicas permite selecionar insetos (os que servem para sua alimentação) em pleno vôo graças a correlatores construídos por aprendizado (adaptação) em camadas de pré-processamento de uma rede neural, que integra várias regiões especializadas do córtex. As características extraídas (*feature extraction*) são a velocidade do batimento das asas (por efeito Doppler) e o tamanho. É um bom exemplo biológico de sistema adaptativo baseado em redes neurais.

A informação que caracteriza um gesto, de interesse deste trabalho, ou que o diferencia de outros gestos, é o seu formato. É este então o conteúdo de informação que se quer registrar fielmente na amostra de um gesto completo e posteriormente treinar a rede neural para reconhecê-lo. O procedimento de pré-processamento deve então garantir que dois gestos de formatos idênticos tenham a mesma representação, mesmo que ocorram diferenças no conjunto de dados obtidos pelo circuito eletrônico. Dois gestos idênticos podem ser executados com dinâmicas diferentes, de forma que as amostras de posição em três dimensões tomadas a cada 12,8 milissegundos (veja cap. 2) , ou seja, em intervalos de tempo uniformes, resultarão em representações diferentes. Para evitar este problema, os gestos passam por uma reamostragem, para uniformizar os espaços entre amostras, e diminuir, se conveniente, o número de amostras para um mínimo sem que se perca informações importantes dos gestos (curvas de pequeno raio, que diferencie um gesto de outro, por exemplo, entre o C e o G)

A reamostragem tem início pelo cálculo do comprimento padrão estimado para os segmentos de reta que formarão a representação final do gesto. Isto é feito dividindo o comprimento total do gesto original (128 segmentos) pelo número de segmentos desejado (32, por exemplo). A partir daí, algumas tentativas serão feitas recursivamente para reajustar este comprimento padrão, de forma a chegar ao número de segmentos igual ao desejado em uma varredura, do primeiro até o último ponto original. Cada tentativa consiste em varrer a seqüência de pontos original, iniciando por fixar o primeiro ponto (que permanece o mesmo) como origem, procurando entre os segmentos subsequentes aquele que contém um ponto cuja distância à origem é igual à distância padrão corrente (figuras 3.6 e 3.7). Encontrado este segmento, é encontrado o ponto de interesse por interseção da equação deste segmento com a circunferência com centro na origem e raio igual à distância padrão.

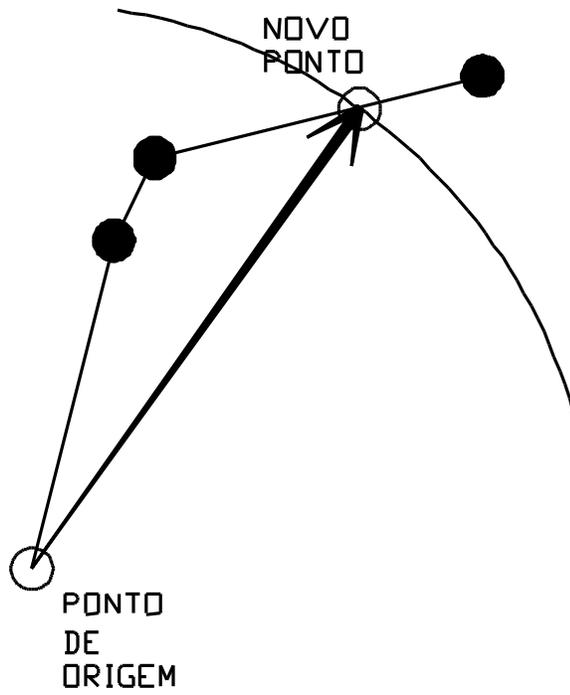


Figura 3.6 – Reamostragem – do 1º ao 2º pontos

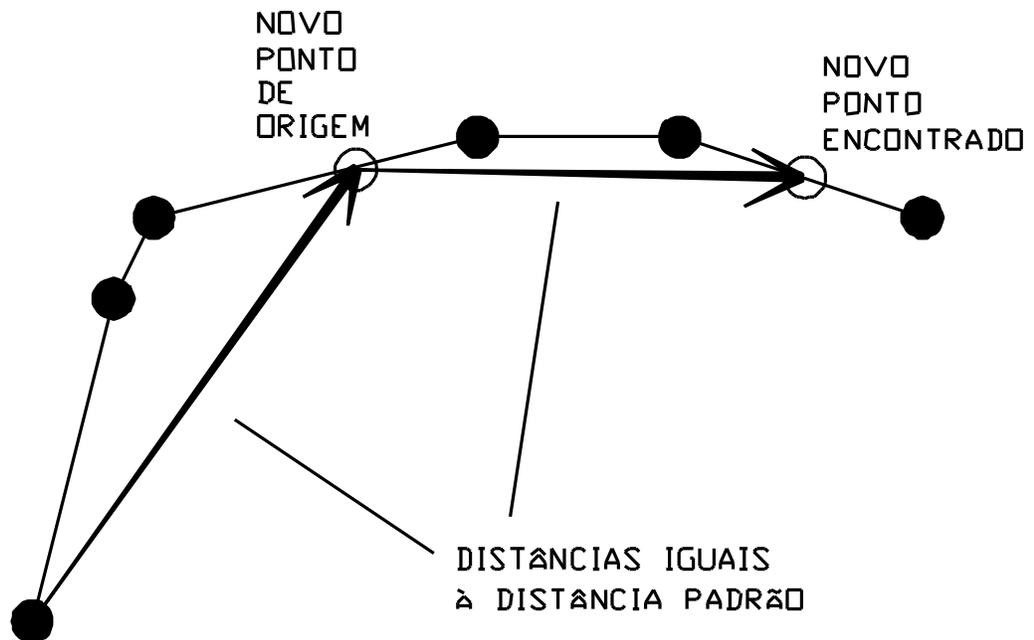


Figura 3.7 – Reamostragem – do 2º ao 3º pontos

O ponto assim encontrado é fixado como parte do gesto reamostrado e será a nova origem para a busca do próximo ponto. Assim é feito sucessivamente até o último ponto original, quando é avaliado se o número de pontos é igual ao desejado; se não é, a distância padrão é recalculada proporcional à diferença do número de pontos alcançado na última tentativa para o desejado.

$$dpr = dpa \times \frac{npc}{npd}$$

onde: dpr = distância padrão reajustada
 dpa = distância padrão anterior
 npc = número de pontos corrente
 npd = número de pontos desejado

A representação de um gesto, neste ponto, é uma seqüência de segmentos de reta cujos comprimentos são idênticos entre si (normalizados, ou seja, todos com módulo=1) . Para cada segmento, é calculado o seu ângulo de inclinação (de $-\pi$ a $+\pi$) sem ambigüidade entre quadrantes 1 e 3, ou 2 e 4, por serem conhecidas as orientações dos segmentos devido à ordenação dos pontos de um gesto em seqüência.

O arquivo de amostras que representam um gesto completo passa a conter os ângulos de inclinação de um número conveniente de segmentos (32, por exemplo), igualmente espaçados entre si. O procedimento descrito já providencia a adequação do formato dos dados à entrada da rede neural, pois pequenos valores de ângulos são muito mais negligenciáveis que pequenos valores absolutos de uma das coordenadas (se os valores absolutos das outras coordenadas também forem pequenos). O número de dados (ângulos) para um gesto pode ser qualquer um entre 1 e 128, mas foram feitos testes com 64, 48, 32, 16, 8.

Nas figuras 3.8 e 3.9 se vêem os resultados obtido com a reamostragem de gestos correspondentes às letras “C” e “Z” , com o número de pontos desejados igual a 32. Os formatos dos gestos foram visivelmente mantidos, e a amostragem foi uniformizada (veja nos gestos originais as variações de comprimentos dos segmentos que correspondem às variações de velocidade durante a execução dos gestos) . A normalização é obtida posteriormente pelo simples fato de que são gravados os valores dos ângulos de inclinação dos segmentos vistos como vetores, na forma da figura anterior. De posse apenas dos ângulos, os comprimentos dos segmentos podem assumir qualquer valor, incluindo a unidade.

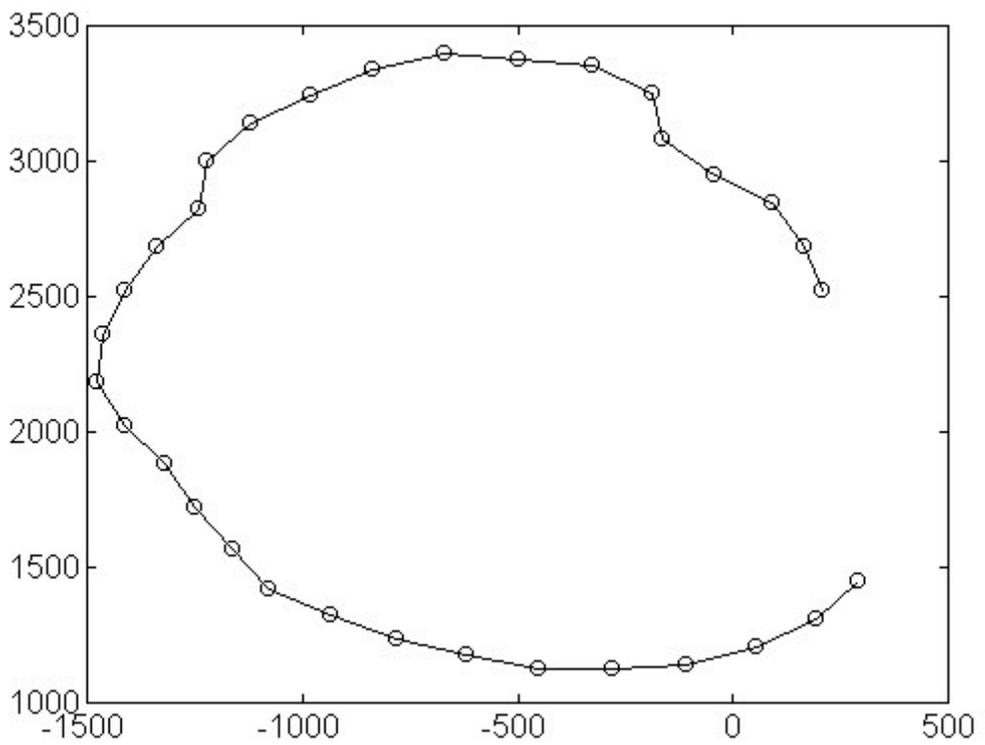
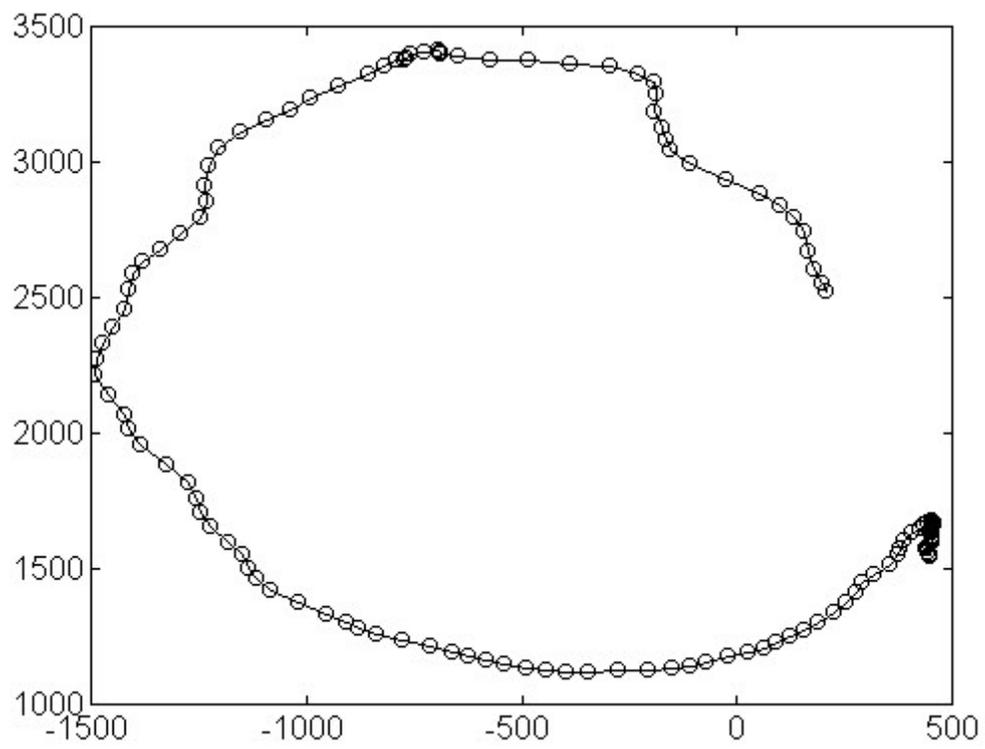


Figura 3.8 – Reamostragem – Letra C

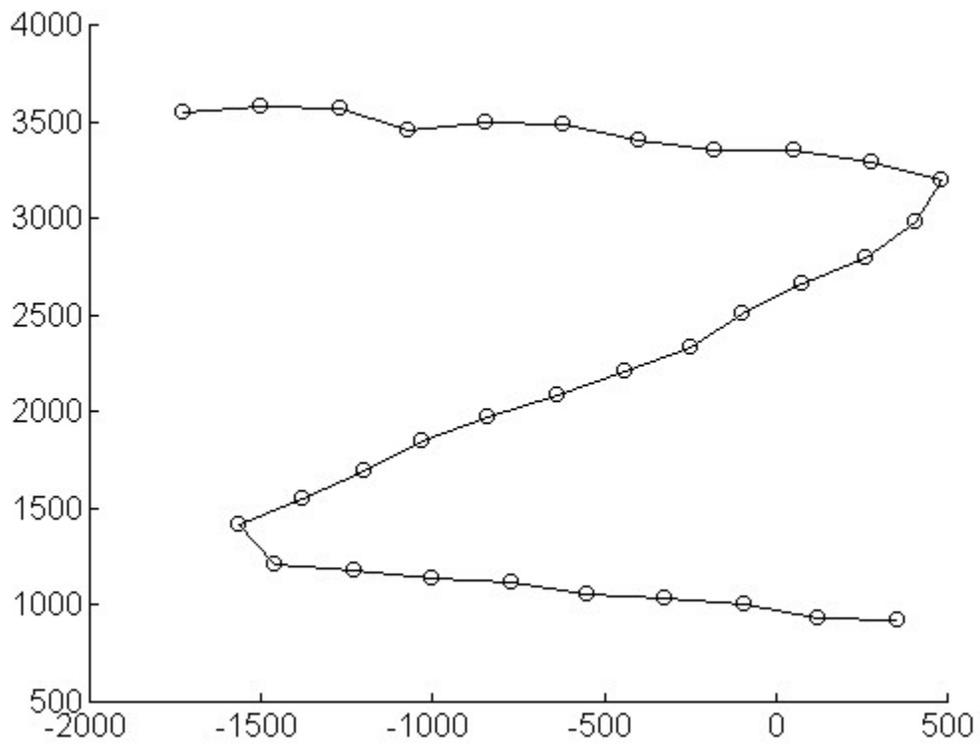
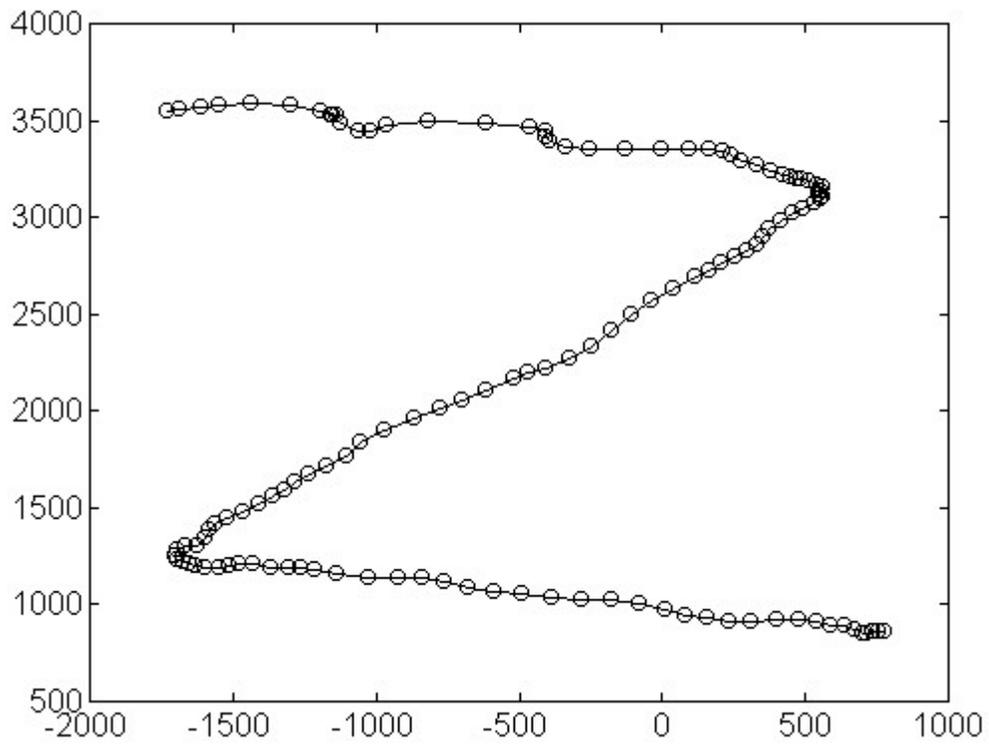


Figura 3.9 – Reamostragem – Letra Z

Para a apresentação dos dados à rede neural para treinamento e avaliação, 40 gestos de cada um dos 8 tipos (a, c, J, n, r, w, y, Z) foram agrupados em um arquivo que contém uma matriz de 320 colunas (40 x 8) e um número de linhas igual ao número de amostras de um gesto (320 colunas x 64 linhas para gestos com 64 amostras = ângulos). Dos 40 gestos, a maior parte é usada para treinamento, uma pequena parte para testar se a convergência é feita com generalização, e outra também pequena para avaliação final do grau de generalização para dados diferentes dos usados no treinamento e nos testes de convergência. Este procedimento tem por objetivo evitar que a rede se especialize nos primeiros gestos (de treinamento), usando como critério de desempenho quantos gestos são reconhecidos em conjuntos diferentes do primeiro [28,31,34].

O número de entradas da rede neural é igual ao número de amostras de um gesto, portanto a diminuição do número de amostras influi diretamente na complexidade da mesma, e no custo computacional de treinamento e posteriormente na velocidade de processamento usando os parâmetros obtidos no treinamento.

A forma de apresentação de cada amostra de uma seqüência a uma entrada da rede neural só é válida quando se pode treinar a rede para um conjunto de dados e confiar que ele reflita a estatística de quaisquer dados que se espera ter nas entradas. Para os sistemas em que esta estatística muda com o tempo (denominados não estacionários), devem ser usadas as topologias de redes dinâmicas, mencionadas no item 3.3, que incluem informações de variação das grandezas de entradas e/ou saídas com o tempo [28,31].

3.4 Métodos de Treinamento, Topologias Usadas e Resultados

O treinamento de uma rede neural do tipo Perceptron Multicamadas pode ser feito na sua forma mais tradicional, com o algoritmo de *backpropagation*, ou incluir algumas sofisticções como taxa de aprendizado adaptativa e momento de inércia, bastante conhecidas e testadas nos mais diversos tipos de aplicações [28,31,34,60], que podem acelerar a convergência para a solução aceitável. A mesma estrutura pode ser usada com algoritmos bem diferenciados na otimização da velocidade de treinamento e qualidade dos resultados, como por exemplo o trabalho em [25,60], onde é usado um métodos de otimização não linear. Os resultados são notavelmente melhores, mas o uso de grande quantidade de memória para cálculos intermediários tornam seu uso, em muitos casos, inviável.

O algoritmo de *backpropagation* é um tipo de aprendizado supervisionado: são apresentados exemplos de dados de entradas associados com as saídas desejadas correspondentes; para cada amostra de dados nas entradas, a rede faz associações entre as mesmas usando os parâmetros correntes (pesos sinápticos) avaliando os valores das saídas; as diferenças entre os valores de saída obtidos e os desejados são usados para calcular o quanto será modificado em cada peso sináptico (inicializados de forma aleatória) de cada neurônio, em direção a minimizar estas diferenças.

Neste trabalho foram usados os algoritmos de *backpropagation* tradicional e outro com taxa de aprendizado adaptativa e momento de inércia. O tradicional foi implementado usando linguagem C e adaptado para execução em Matlab, para tornar mais prática a apresentação de resultados. Como os resultados iniciais não eram satisfatórios, foi modificado o algoritmo com taxa de aprendizado adaptativa e momento de inércia disponível em bibliotecas do Matlab [33], acrescentando a validação periódica com exemplos diferentes dos usados treinamento. Após as correções feitas na forma de apresentação dos dados de entrada (reamostragem para comprimentos uniformes), foram obtidos resultados melhores que os esperados, neste procedimento resumido de exemplificação de uso da interface gestual. Foi repetido o teste do programa de *backpropagation* em C / Matlab com a forma corrigida de apresentação de dados, que indicaram boas perspectivas para sua utilização com o acréscimo no futuro de otimizações de velocidade de execução (compilação em C, acréscimo de taxa de aprendizado adaptativa e momento de inércia). Várias simulações foram feitas para encontrar um dimensionamento satisfatório das camadas intermediárias, sempre nas mesmas condições de teste. Foram testes variadas também as proporções entre os conjuntos de dados para treinamento-teste-validação, como 30-5-5, 20-10-10, 25-10-5, com melhores perspectivas para a primeira opção. Os resultados mais esclarecedores usando a melhor configuração são mostrados em seguida, acrescidos de impressões sobre as perspectivas de melhorias dos resultados.

3.4.1 Resultados de simulações e interpretações:

Após terem sido testadas várias configurações, com várias simulações para reconhecer os comportamentos médios, foi escolhida aquela que reunia bons resultados sem exageros nas dimensões e no tempo de processamento. Para os oito testes relatados em seguida, a configuração usada foi: 64 entradas, 1ª camada escondida com 32 neurônios, 2ª camada escondida de 16 neurônios, camada de saída com 8 neurônios. Foram usadas 40 amostras de gestos de 64 pontos (após reamostragem), sendo que 30 participaram da fase de treinamento, 5 foram usados para teste de generalização, 5 na validação da solução final. O erro quadrático médio aceitável foi estipulado em 0,01. Em alguns testes foram usados gestos amostrados por dois usuários diferentes, para experimentar se o comportamento de uma rede treinada se mantém satisfatório ao trocar de usuário sem repetir o treinamento. O resumo dos resultados e as médias de desempenho para os dados de validação foram:

TESTE	a	c	j	n	r	w	y	z	MÉDIA %
1	100	80	100	100	100	100	100	100	98
2	80	80	80	100	100	100	100	60	88
3	80	80	80	100	100	100	100	60	88
4	80	90	80	100	100	100	100	90	93
5	90	80	90	100	100	100	100	85	93
6	100	80	100	100	100	100	100	100	98
7	80	90	80	100	100	100	100	90	93
8	100	80	80	80	100	100	100	100	93
MÉDIA %	89	83	86	98	100	100	100	86	93

Observações feitas em cada teste:

- teste 1 - chegou em $ea=0.01$ (erro aceitável) em alguns minutos;
- teste 2 - repetição do teste 1, para confirmá-la como a melhor configuração: também chegou em $ea=0.01$ em alguns minutos, mas com desempenho ruim em "z";
- teste 3 - continuação do teste 2 (continuando o treinamento com critérios mais apertados), para confirmá-la como a melhor configuração: chegou em $ea=0.001$ em alguns minutos, mas com desempenho ruim em "z" ; não melhorou nada em comparação com $ea=0.01$; continuando a treinar até atingir o critério $ea=0.0001$, também não houve melhorias; a conclusão: o que já se alcança em termos de generalização (desempenho com amostras de validação) com $ea=0.01$, não vai melhorar muito chegando a menores erros nas amostras de teste; a vantagem em comparação com o teste 1 é que não há gestos reconhecidos em duas classes;
- teste 4 - nas mesmas condições do teste 1, mas encadeando os gestos de dois usuários nas matrizes de amostras; apesar de muitas das letras terem resultado em 100% de reconhecimento, um gesto da letra "n" foi também reconhecida como "a";
- teste 5 - repetição do teste 4, para confirmar o desempenho do mesmo; apesar de vários 100%, algumas das letras foram reconhecidas como outras, como no teste 4;
- teste 6 - repetição do teste 4, uma amostra "c" é classificada como "a" e uma "j" como "j" e "a" ao mesmo tempo;
- teste 7 - repetição do teste 4, com dois usuários, resultados semelhantes;
- teste 8 - repetição do teste 4, com dois usuários, resultados semelhantes.

Os gráficos a seguir (figuras 3.10 e 3.11) representam a classificação dos gestos amostrados para validação do aprendizado. O primeiro gráfico tem no eixo horizontal a numeração dos gestos amostrados, separados de grupos 5 pertencentes ao mesmo gesto pretendido (de 1 a 5 para a letra "a", de 6 a 10 para a letra "c", e assim por diante na mesma ordem da tabela anterior). O eixo vertical traz os neurônios cujas saídas foram ativadas quando cada gesto foi apresentado às entradas da rede treinada (cada neurônio corresponde a uma classe). Temos então a relação entrada-saída para uma rede treinada: o exemplo visto teve todos os gestos reconhecidos em suas classes corretas, mas o gesto nº 20, da classe 4 (um "n"), foi reconhecido erroneamente como sendo também da classe 1 (da letra "a").

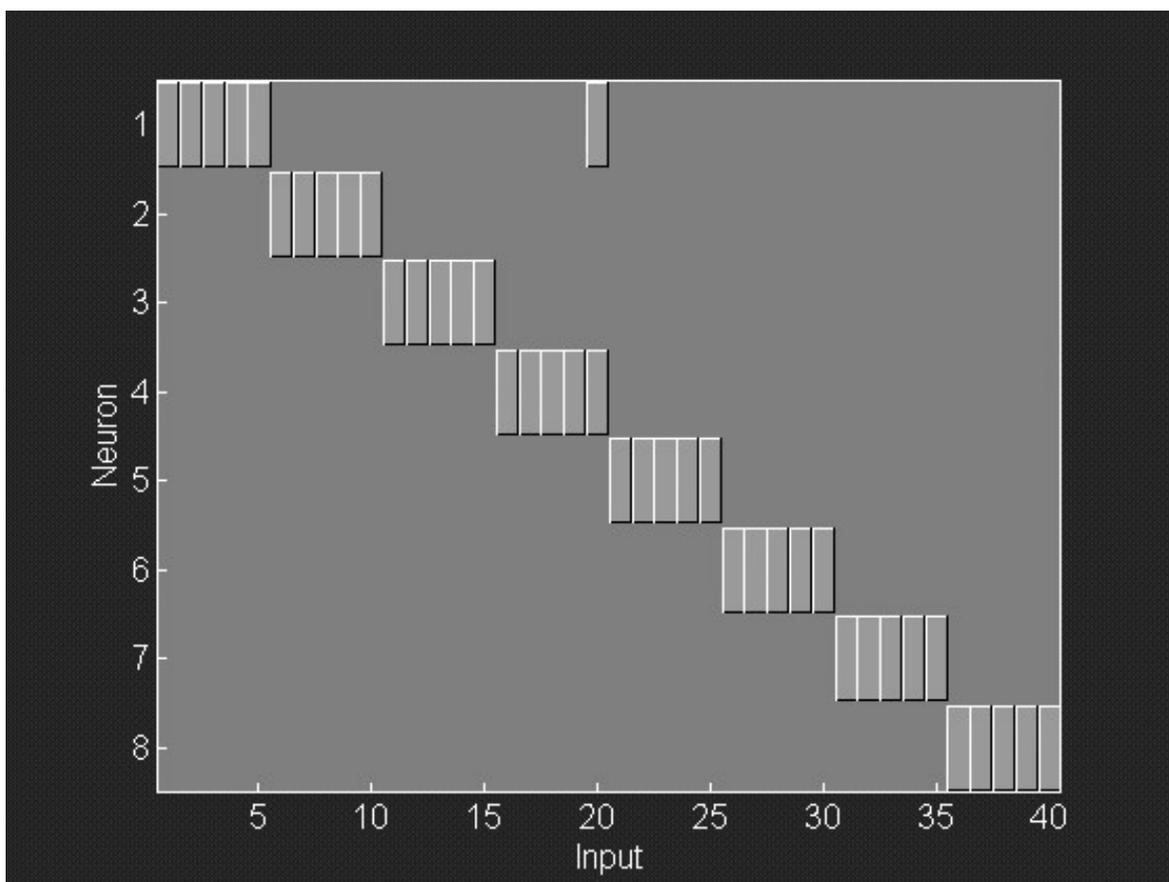


Figura 3.10 – Classificação dos gestos amostrados – 1 usuário

O segundo gráfico mostra a apresentação simultânea dos gestos de dois usuários (1 a 40 para o primeiro, 41 a 80 para o segundo), em fase de validação, tal como foi feita no treinamento. No exemplo escolhido aparecem outros tipos de erros de classificação: o gesto nº 39 não foi reconhecido como sendo da classe 8, e de nenhuma outra; o gesto 7 (“c”) não foi reconhecido em sua classe esperada, mas sim na classe 5 (“r”). Os gestos de nºs 32 e 35 foram reconhecidos na sua classe correta (classe 7, dos “y”) e também na classe 6 (“w”) de forma incorreta, como no exemplo do gráfico anterior.

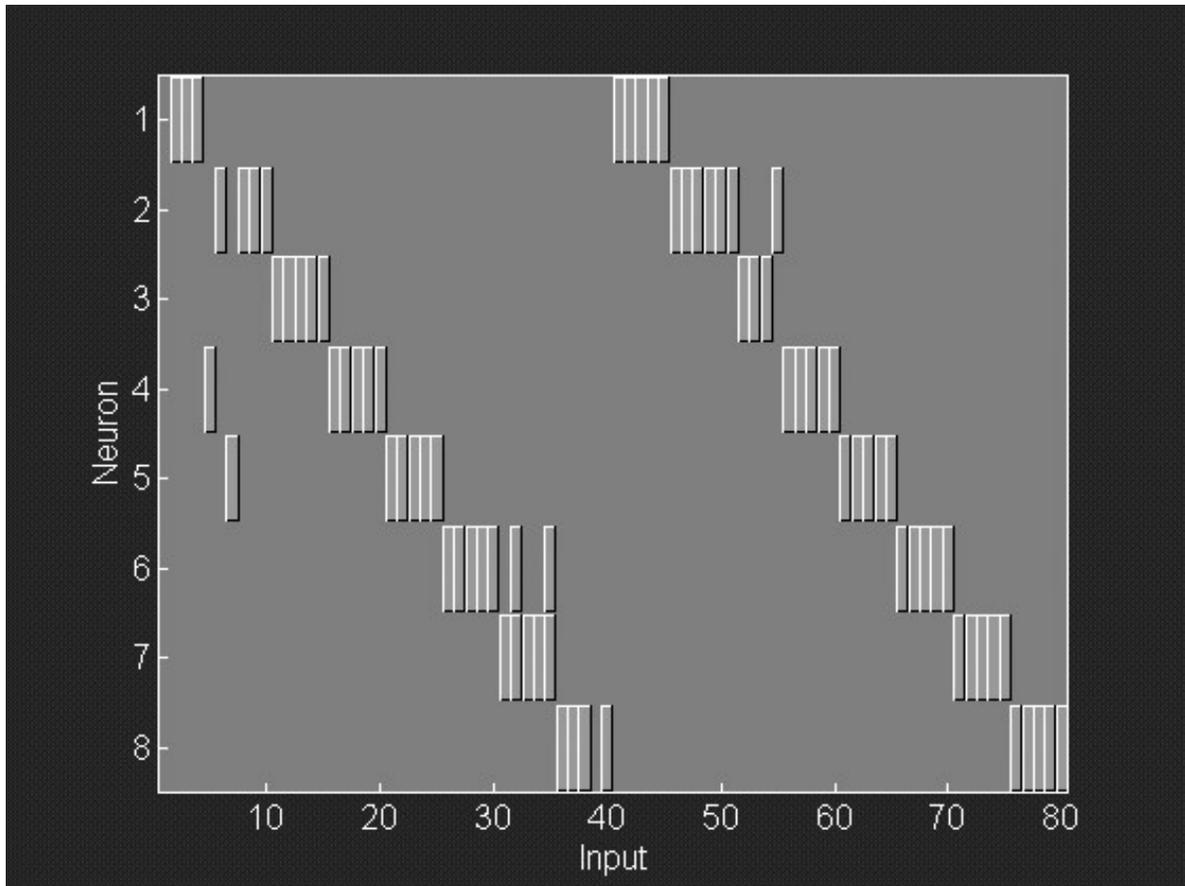


Figura 3.11 – Classificação dos gestos amostrados – 2 usuários

Nos testes feitos com outras dimensões nas camadas intermediárias, se destacou a configuração com 64 entradas - 16 neurônios escondidos - 8 neurônios escondidos - 8 neurônios na saída, pois nesta, com menos neurônios, mas dispendo de mais tempo de treinamento para alcançar menores índices de erros no conjunto de dados de teste, se alcança melhores resultados com os dados de validação. Dois dos testes têm seus resultados agrupados na tabela a seguir, para ilustrar as observações feitas.

TESTE	a	c	j	n	r	w	y	z	MÉDIA
9	80	100	80	100	100	100	80	60	88
10	80	100	100	100	100	100	80	60	90
MÉDIA	80	100	90	100	100	100	80	60	<u>89</u>

Observações :

- a) Do teste 9 ($ea=0.01$) para o 10 ($ea=0.003$) só melhorou o desempenho com a letra "j", de 80 para 100 %; este resultado apareceu em vários outros testes que não foram registrados .
- b) Há uma grande vantagem se comparada c/ a rede dos testes 1 a 8: não há gestos reconhecidos em 2 classes simultaneamente, indicando que reduzir as dimensões das camadas escondidas de neurônios tem alguma influência na melhoria da separação de classes (funções mais simples gerando superfícies de separação menos detalhadas), quando o objetivo é favorecer a generalização [28,31,34] ; este resultado também se repetiu nos testes não registrados .

Observações gerais:

Assim como em outras simulações de redes neurais deste tipo, mesmo na solução de problemas bem diferentes deste, foi observado um limite na capacidade de evoluir para soluções melhores, ou menores erros quadráticos, usando repetidamente um determinado conjunto de dados de treinamento e variando a topologia. Em outras palavras, no estudo da solução de um problema por simulações, progressivamente aumento o número de neurônios por camada e o número de camadas; quando encontro uma certa dimensão para a rede neural a partir da qual as soluções não mostram quase nenhuma evolução, atingi o limite citado. O que se observa a partir daí, é que um aumento de dimensões permite atingir uma solução parecida com a anterior, por meio de um número menor de ciclos de treinamento, mas um ciclo demora mais por haver mais neurônios, ficando o tempo total de conversão, por fim, bem próximo do tempo gasto na configuração anterior (e na configuração em que se atingiu o citado limite). Parece que uma rede neural alcança, no máximo, uma solução que corresponde à complexidade (ou variedade) das informações contidas no conjunto de dados. Simulações com outros conjuntos de dados atingem limites diferentes.

A diminuição do número de pontos em um gesto para 32, como resultado de reamostragem, com a conseqüente diminuição do número de entradas da rede neural, faz diminuir o tempo de processamento, mas faz piorarem os índices de reconhecimento dos gestos. Parece que com 32 pontos, detalhes dos gestos que são importantes na sua diferenciação são negligenciados.

Os erros nos conjuntos de dados de validação são maiores para as redes neurais com apenas uma camada intermediária, o que se deve ao fato de as camadas intermediárias sucessivas se prestarem a diferenciar características locais de trechos dos gestos, essencial na identificação de superfícies complexas de separação entre as classes. A complexidade não pode ser exagerada, por outro lado, para favorecer a generalização do aprendizado, principalmente quando se tem um número reduzido de dados de treinamento. Alguns pesquisadores treinam a rede até que os erros sejam várias ordens de magnitude abaixo dos valores manipulados, sem computar os erros em dados diferentes (dados de testes) especializando a solução para os dados de treinamento, o que piora também o aspecto de

generalização. Neste trabalho, bons resultados foram obtidos tomando como critério de finalização do treinamento a minimização de erros em um conjunto de dados de testes que não participam do treinamento.

O fato de ter sido atingido índices de acerto por volta de 90 % nos dados de validação indica que a rede neural conseguirá reconhecer bem os gestos feitos após a fase de treinamento. O que não está explicitado nos dados das tabelas apresentadas (mas aparece nos gráficos) é a quantidade de gestos que são ditos como pertencentes a classes diferentes da pretendida. Neste aspecto, foi observado que o índice de classificações erradas é bastante pequeno. Pode-se dizer então que foram atingidos bons resultados em termos de reconhecimento e de separação de classes de gestos.

Capítulo 4 - Conclusões e Perspectivas Futuras

4.1 Sobre a Medida de Posição por Ultra-Som

Algumas conclusões puderam ser tiradas sobre a medida de posição por ultra-som, e outras não experimentadas podem ser intuídas a partir deste trabalho e dos textos anteriormente publicados sobre o assunto. A necessidade de um equilíbrio entre a simplicidade pretendida e uma precisão aceitável remeteu ao adiamento de algumas questões sobre a avaliação formal da exatidão alcançável nas medidas, em termos absolutos, mas permitiu concentrar as atenções na avaliação de resultados práticos do trabalho aplicados a um campo de grande interesse na evolução da forma de uso das máquinas : a comunicação por meio de gestos.

A escolha por colocar um transmissor de ultra-som no objeto a ser localizado causou algum incômodo no uso da ferramenta para captura de gestos, por ser muito direcional o emissor de ultra-som usado, limitando os gestos aos que são possíveis com o dedo indicador apontando para os sensores. Para continuar trabalhando com a mesma configuração de sensores, o emissor poderia ser substituído por uma montagem de alguns emissores dispostos como se estivessem fixados em uma seção esférica, conjugando suas curvas de atenuação espacial de forma a compor um ângulo de emissão maior.

A forma de cálculo das coordenadas retangulares em três dimensões a partir das informações de distâncias do transmissor até três sensores coplanares poderia também causar limitações nos movimentos do usuário devido à limitação do ângulo de resposta dos sensores, pois estes são apontados para a mesma direção, ortogonal ao plano formado por eles. O efeito de incômodo ligado a este fator é muito inferior ao causado pelo limite de amplitude angular do transmissor, desde que se mantenha uma distância razoável entre sensor-transmissor (acima de 20 cm). Isso se explica por dois motivos:

1- quanto maior for a distância entre o transmissor e os sensores, menor será o ângulo visto pelos transmissores se compararmos movimentos de mesmo comprimento.

2- é muito mais fácil vigiar para que a amplitude do gesto seja pequena o suficiente do que cuidar para que o dedo indicador aponte sempre na mesma direção.

Foi observado que, enquanto os sensores se encontravam nos vértices de um triângulo equilátero de 70 mm de lado, havia uma precisão limitada nos valores calculados, em grande parte resolvida pelo afastamento dos sensores para os vértices de um triângulo equilátero de 203 mm de lado. Esta limitação advém da diferença, em ordens de grandeza, entre as distâncias medidas e as diferenças entre elas, como explicado no item 2.3. Se a natureza física da medida de distâncias por ultra-som a 40 kHz limita a resolução espacial em 4,25 mm (veja item 2.5), as diferenças entre as medidas de distância também poderiam aparecer somadas com erros inaceitáveis. Na prática, como já foi dito antes, este limite não inclui as parcelas de

incerteza da medição, e o resultado final obtido incluindo o pré-processamento foi melhor do que o previsto.

Havia uma idéia de se usar, para o mesmo objetivo deste trabalho, uma matriz com 5 colunas por 3 linhas de sensores, com transmissão de pulsos e recepção de ecos com padrões de ganhos direcionais adaptáveis (*beamforming*), usando um DSP (veja item 2.2.5) para implementar algoritmos de filtros digitais, mas o critério de baixo custo levou a uma reavaliação da estratégia de interface com o usuário e das necessidades de processamento, tornando mais conveniente se partir para a simplificação da solução. Esta outra proposta possibilitaria a extração de informações mais completas, como a posição e a dinâmica dos dedos da mão em movimento.

Durante a avaliação e o ajuste do circuito eletrônico, foi possível experimentar o desempenho de um sensor individual com sinais refletidos pela mão de um usuário, para se ter uma idéia sobre o desempenho do método de pulso-eco. Os níveis de sinal observados são, como visto na literatura [3,53], muito inferiores ao sinal direto, tornando inadequado o uso do circuito de baixo custo empregado, exigindo componentes e topologias mais sensíveis, menos susceptíveis a ruídos e o emprego de filtros digitais para a rejeição de reflexões indesejadas e interferência de ruídos do ambiente, o que aumentaria a complexidade para muito além da pretendida.

Ao longo do desenvolvimento deste trabalho, tanto o autor quanto as pessoas que tinham algum conhecimento do mesmo foram vislumbrando algumas extensões, das quais julgo serem possíveis com o emprego de ultra-som:

1- o uso de medidas por transmissão-reflexão, que livraria o objeto a ser localizado de conter um transmissor.

2- o emprego de disposições diferentes para os sensores, para facilitar os cálculos geométricos.

3- o emprego de disposições diferentes para os sensores, para possibilitar o emprego de filtros digitais e facilitar a implementação do item 1.

4- o emprego de modulação do sinal de ultra-som (como é feito em controles remotos de televisão) para codificação que diferenciasse mais de um objeto, aumentando a utilidade do sistema criado.

5- a experimentação com transmissores/sensores menos direcionais e mais potentes/sensíveis, com outras frequências de trabalho, feitos sob medida para esta aplicação.

6- a disponibilização de uma saída direta em tensão ou corrente, em escalas padronizadas para instrumentação, para cada coordenada, e outras alterações que possibilitam seu uso em ambiente industrial.

7- a adequação do circuito para uso no estudo detalhado do movimento individual de cada articulação do corpo humano, ou para estudo de outros movimentos em três dimensões.

8- a adaptação do circuito e de *software* para servir como dispositivo de entrada para instrumentos musicais virtuais, no qual as notas pudessem ser escolhidas no espaço, com uma dose extra de expressividade.

9- O uso de onda contínua de ultra-som, com interrupções curtas e cadenciadas (funcionamento ao contrário do que foi implementado) com o objetivo de se ter sempre um

sinal ao qual se adapte um controle automático de ganho nos primeiros estágios de amplificação do sinal.

10- A adaptação do meio de comunicação com o computador para serial RS-232C, e o desenvolvimento de um driver para que a interface criada possa ser vista pelos aplicativos já existentes como um mouse (sem fio e sem necessidade de ser arrastado em uma mesa).

11- Outras idéias possíveis, mais afastadas da linha principal deste trabalho, vindas de saudáveis excursões da imaginação das pessoas.

4.1.1 Perspectivas de Utilização por Músicos

O sistema de interface ultrassônica será usado em um projeto interdisciplinar na Unicamp para o estudo do uso de gestos na a intervenção humana em composição automatizada de música (por computador). A intervenção se dá pelo deslocamento da mão do usuário (cuja posição é amostrada pela interface) em ambientes gráficos virtuais, permitindo a atuação da criatividade humana, de forma intuitiva, nestes algoritmos de composição, que se tornam assim interativos e de tempo real.

Um outro conjunto de aplicações pode ser cogitado a partir do que se vem fazendo com um equipamento chamado Soundbeam (acessível via Home Page da Internet). A maioria das aplicações está ligada à expressão musical sem a necessidade do aprendizado de técnicas, e tem sido usado por escolas para deficientes como valiosa ferramenta pedagógica. Ele usa um transmissor e um sensor de ultrassom, medindo continuamente a distância até um anteparo (por exemplo, a mão) que se coloque à sua frente pelo eco de ultrassom e usa esta informação para controlar parâmetros de geradores de som (ou de iluminação). O sistema de medição de distância da Interface Gestual Ultrassônica é outro (não usa eco), mas fornece a posição em três dimensões, aumentando a possibilidade de controle intuitivo de vários parâmetros ao mesmo tempo, em tempo real.

4.2 Sobre o Reconhecimento de Gestos

O usuário que executou gestos mais livremente, sem tomar muito cuidado com o limite de tempo para um gesto, gerou resultados mais distorcidos, causando uma impressão de que teria desempenho pior no seu reconhecimento. O resultado foi o inverso, pois o critério de fim de treinamento favorece a generalização, e uma maior variância nos dados de treinamento parece ajudar neste aspecto.

Na rede neural, foi usado um limiar fixo, para tomada de decisão da camada de neurônios de saída, igual a 50% (decisão por ser um gesto pertencente a uma determinada classe). Para melhorar o desempenho geral do reconhecimento de gestos e separação de classes, é melhor que tenha um limiar variável, igual à mediana do conjunto dos valores

amostrados, ou tomando por vencedora uma classe com vantagem percentual mínima sobre as outras (30%, por exemplo) . O reconhecimento (errôneo) de um gesto como dois comandos possíveis poderia ser resolvido com um limiar de separação superior e, por outro lado, para diminuir o número de gestos não reconhecidos seria usado um limiar mais baixo

Um gesto feito pelo usuário é amostrado como uma série de pontos encadeados no percurso descrito pelo transmissor (no dedo do usuário). Um conjunto de 128 pontos (aproximadamente 1 segundo e meio de amostragem) é considerado um gesto completo. Outras formas de identificar a separação entre os gestos a serem considerados e os traços percorridos antes e depois de cada gesto são sugeridas no cap. 3, item 3.1. A evolução deste aspecto é essencial para tornar possível o uso dessa interface no auxílio aos desabilitados, que podem exigir cuidados ainda maiores, por fazerem movimentos mais tremidos, ou de menores amplitudes, ou com distorções ligadas aos problemas específicos de mobilidade.

Usando somente a interface de avaliação de posição em 3-D, pode-se estudar novas formas de interface com o usuário, como por exemplos, “put it there” (veja cap. 1), ou um mouse 3-D que opere em “layers” acessados pela variação da proximidade do transmissor aos sensores . Ela também permite ao desenvolvimento de *softwares* que avaliem o reconhecimento de gestos por outros métodos, como as Cadeias de Markov, muito usadas na classificação de seqüências de fonemas nos sistemas de reconhecimento de fala. Um outro possível uso como *feedback* para sistemas artificiais de coordenação muscular permitiria o detalhamento de percursos das articulações em movimento.

Observação Final

Os objetivos iniciais deste trabalho foram plenamente alcançados, tanto no que diz respeito às implementações em *hardware* e *software*, quanto na criação de perspectivas para trabalhos futuros. A opção por um trabalho que incluísse a implementação em *hardware*, indo além das simulações, foi feita por se adequar ao perfil do autor deste trabalho e da possibilidade de fornecer ferramentas para outros trabalhos de pesquisa na Universidade.

Apêndice A - Listagem do Programa de Aquisição de Gestos

```
/* #####                                     ##### */
/* #####                                     ##### */
/* #####          GI22.C - Controle de versões          ##### */
/* #####                                     ##### */
/* ##### - Filtra os raios usados no calculo de coordenadas xyz do ##### */
/* ##### transmissor (media entre ponto atual e ponto anterior) ##### */
/* #####                                     ##### */
/* ##### - Adaptado ao aumento da distancia entre sensores p/ 20,3cm ##### */
/* #####                                     ##### */
/* ##### - Gera uma sequencia de arquivos com amostras de um mesmo ##### */
/* ##### gesto de um mesmo usuario                    ##### */
/* #####                                     ##### */
/* ##### - Passa a gravar nos arquivos de gestos somente X, Y, Z ##### */
/* #####                                     ##### */
/* #####                                     ##### */

#include <stdio.h>          /* Contem o prototipo para printf          */
#include <dos.h>            /* Contem o prototipo para outportb e inportb */
#include <conio.h>         /* Contem o prototipo para getch e clrscr    */
#include <math.h>          /* Contem o prototipo para pow              */
#include <string.h>        /* Contem o prototipo para strepy          */
#include <stdlib.h>        /* Contem o prototipo para getenv e exit    */

#define ctrl_mask 0x04    /* Seta bits 0 a 3 Cntrl Port => Open Col.  */
#define INT_No 0x0F      /* No. do vetor p/ INT7                      */
#define npontos 0x80     /* Numero de pontos de um gesto             */

void main(void);
void getfile(void);
void interrupt far readisr(void);          /* Interr. Service Routine(ISR)            */
void interrupt far (*oldhandler)();        /* Antigo vetor da INT7                    */
void exit(int status);
unsigned char lenibble(unsigned char end_nibble);
int numera_outfile(void);

int ocorreu_int, st_read, status_read, irq_ativa;
unsigned char data_read;
unsigned int nfile, maxfiles;;
char nomebase[10] ;
char *continua ;
int d;
```

```

void interrupt far readisr(void)                /* Interr. Service Routine(ISR) */
{
    disable();

    /* ##### Inicio da Rotina de Servico a Interrupcao - ISR ### */

    status_read = (inportb(0x379) & 0x40);
    if (status_read && irq_ativa)
    {
        data_read = inportb(0x378);            /* le barram. Dados */

        outportb(0x378, (data_read & 0xFD)); /* 8+ = escr. "0" em STOPEN */
        delay(d);                             /* +0 = "0" em BD1(/STROBE) */
        outportb(0x378, (data_read & 0xFD)); /* Repete "0" por garantia */
        delay(d);
        outportb(0x378, (data_read | 0x02)); /* Nivel "1" em BD1 */
        delay(d);
        outportb(0x378, (data_read & 0xFD)); /* Nivel "0" em BD1 */
        delay(d);

        ocorreu_int = 1;

        /* Escr. "0" em STOPEN p/ inibir /INTR e preparar p/prox. STARTEN */
        /* Pulso 0-1-0 no gate do lach de entrada da EPLD (BD0) */

        outportb(0x378, 8+0);                 /* 8+ = escr. "0" em STOPEN */
        delay(d);                             /* +0 = "0" em BD0(/STROBE) */
        outportb(0x378, 8+0);                 /* Repete "0" por garantia */
        delay(d);
        outportb(0x378, 8+1);                 /* Nivel "1" em BD0 */
        delay(d);
        outportb(0x378, 8+0);                 /* Nivel "0" em BD0 */
        delay(d);
    }

    /* ##### Fim da Rotina de Servico a Interrupcao - ISR ### */

    /* oldhandler();                          Passa controle p/ antiga ISR */
    outportb(0x20,0x20);                       /* Envia EOI ao PIC1 */
    enable();
}

```

```

void main(void)
{

    clrscr();
    printf ("\n  Amostragem de Gestos para Treinamento de Redes Neurais \n ");
    printf ("\n Nome-base para os arquivos = ");
    gets (nomebase);
    printf ("\n          %s\n",nomebase);
    printf ("\n Numero inicial para os arquivos = ");
    scanf ("%d", &nfile);
    printf ("\n Numero final para os arquivos = ");
    scanf ("%d", &maxfiles);
    printf ("\n\n");
    gets (continua);

    while (nfile <= maxfiles)
    {

        printf("\n Pressione ENTER para amostrar 1 gesto: [ f + <ENTER> finaliza ]");
        gets (continua);
        if (continua[0] == 'f' || continua[0] == 'F') break ;
        printf("\n Execute 1 gesto : ");

        getfile();                                /* funcao que gera um arquivo xx.dat */

        numera_outfile();
        nfile++;

    }

}

void getfile(void)
{

    FILE *outfile_ptr;
    int npt_atual, cont;
    unsigned char end_atual, nibble, nibble_count, pic21;
    unsigned char control_read, control_write;
    int gesto[npontos][16];
    double word, raio_atual, raio_medio, x, y, z;
    double RQ[npontos][5];
    double R_ant[5];
    double R_atual[5] = {0, 0, 0, 0, 0};

    outfile_ptr = fopen("xx.dat", "w");

```

```

if (outfile_ptr == NULL)
{
    printf("\nNao foi possivel abrir o arquivo .dat");
    exit(1);
}

d=0.1;
ocorreu_int = 0;

/* ##### Determina situacao inicial ##### */

/* Mascara a interrupcao */

pic21 = inportb(0x21);          /* Le mascaras de interrupcao */
outportb(0x21,(pic21 | 0x80)); /* Seta mascara INT7 (PIC) */
pic21 = inportb(0x21);        /* Le mascaras de interrupcao */

/* Escreve "0" no bit 4/Control Port = inibe interr. IRQ7 */

control_read = inportb(0x378+2); /* le control port */
control_write = control_read & 0xEF; /* mascara control port */
outportb(0x378+2, control_write); /* inibe IRQ7/hardware */
delay(d);
outportb(0x378+2, control_write); /* repete para garantir */
delay(d);
control_read = inportb(0x378+2); /* le control port */

/* Escreve "0" em STOPEN para inibir /INTR e preparar p/ prox. STARTEN */
/* Pulso 0-1-0 no gate do lath de entrada da EPLD (BD0) */

data_read = inportb(0x378); /* le barram. dados */
outportb(0x378, 8+0); /* 8+ = escr. "0" em STOPEN */
delay(d); /* +0 = "0" em BD0(/STROBE) */
outportb(0x378, 8+0); /* Repete "0" por garantia */
delay(d);
outportb(0x378, 8+1); /* Nivel "1" em BD0 */
delay(d);
outportb(0x378, 8+0); /* Nivel "0" em BD0 */
delay(d);
data_read = inportb(0x378); /* le barram. dados */

/* Verifica a polaridade da IRQ */

st_read = inportb(0x379); /* le um byte do port de status */
irq_ativa = !(st_read & 0x40);

```

```

/* Prepara para interrupcao */

oldhandler=getvect(INT_No);          /* Salva antigo vetor da INT7 */
setvect(INT_No,readisr);             /* Seta o novo vetor da INT7 */
pic21 = inportb(0x21);               /* Le mascaras de interrupcao */
outportb(0x21,(pic21 & 0x7F));       /* Retira mascara da INT7 (PIC) */
pic21 = inportb(0x21);               /* Le mascaras de interrupcao */

/* Escreve "1" no bit 4/Control Port = permite interr. IRQ7 */

control_read = inportb(0x378+2);     /* le control port */
control_write = control_read | 0x10; /* tira mascara contrl port */
outportb(0x378+2, control_write);   /* permite IRQ7/hardware. */
delay(d);
outportb(0x378+2, control_write);   /* repete para garantir */
delay(d);
control_read = inportb(0x378+2);     /* le control port */

/* ##### Inicio da Rotina Principal ##### */

/* Escreve "1" em STOPEN para permitir /INTR e STARTEN */
/* Pulso 0-1-0 no gate do lach de entrada da EPLD (BD0) */

data_read = inportb(0x378);          /* le barram. dados */
outportb(0x378, 0+0);                /* 0+ = escr. "1" em STOPEN */
delay(d);                             /* +0 = "0" em BD0(/STROBE) */
outportb(0x378, 0+0);                /* Repete "0" por garantia */
delay(d);
outportb(0x378, 0+1);                /* Nivel "1" em BD0 */
delay(d);
outportb(0x378, 0+0);                /* Nivel "0" em BD0 */
delay(d);
data_read = inportb(0x378);          /* le barram. dados */

npt_atual = npontos - 1;

while (npt_atual >= 0)
{
/*   delay(d);                       */
/*   status_read = inportb(0x379);   */

    if (ocorreu_int == 1)
    {

```

```

    ocorreu_int = 0;
    for (cont=16 ; cont >= 1; cont--)
    {
        end_atual = cont-1;
        gesto[npt_atual][end_atual] = lenibble(end_atual);
    }
    npt_atual = npt_atual - 1;

    /* Escreve "1" em STOPEN para permitir /INTR e STARTEN */
    /* Pulso 0-1-0 no gate do lach de entrada da EPLD (BD0) */

    delay(d);
    outputb(0x378, 0);          /* 0 = escr. 1 em STOPEN */
    delay(d);                  /* 0 = em /STROBE da EPLD */
    outputb(0x378, 0);        /* Repete 0 p/ garantia */
    delay(d);
    outputb(0x378, 1);        /* Nivel 1 em BD0 */
    delay(d);
    outputb(0x378, 0);        /* Nivel 0 em BD0 */
    delay(d);

}

}

/* ##### Fim da Rotina Principal ##### */

/* Escreve "0" em STOPEN para inibir /INTR e preparar p/ prox. STARTEN */
/* Pulso 0-1-0 no gate do lach de entrada da EPLD (BD0) */

delay(d);
outputb(0x378, 8+0);          /* 8+ = escr. "0" em STOPEN */
delay(d);                    /* +0 = "0" em BD0(/STROBE) */
outputb(0x378, 8+0);        /* Repete "0" por garantia */
delay(d);
outputb(0x378, 8+1);        /* Nivel "1" em BD0 */
delay(d);
outputb(0x378, 8+0);        /* Nivel "0" em BD0 */
delay(d);

/* Escreve "0" no bit 4/Control Port = inibe interr. IRQ7 */

outputb(0x378+2, inportb(0x378+2) & 0xEF); /* inibe IRQ7/hardware */
delay(d);
outputb(0x378+2, inportb(0x378+2) & 0xEF); /* repete para garantir */
delay(d);

```

```

/* Desfaz a preparacao para interrupcao */

pic21 = inportb(0x21);           /* Le mascaras de interrupcao */
outportb(0x21,(pic21 | 0x80));  /* Seta mascara INT7 (PIC) */
pic21 = inportb(0x21);         /* Le mascaras de interrupcao */
setvect(INT_No,oldhandler);    /* Restaura antigo vetor INT7 */

/* Transfere a matriz de dados de medida de um gesto para arq. em disco */

npt_atual = npontos - 2;        /* -2 no lugar de -1 para */
word = 0;                       /* desprezar a 1a medida */

while (npt_atual >= 0)
{
    for (cont=16 ; cont >= 1; cont--)
    {
        end_atual = cont - 1;
        nibble_count = end_atual % 4;

        nibble = gesto [npt_atual] [end_atual];
        word = word + ( nibble * pow(16,nibble_count) );

        if (nibble_count == 0)
        {
            raio_atual = word * 1.7 ;
            /* 1.7=340(m/s)/2MHz*10000 => 0.1mm */
            R_ant[end_atual/4] = R_atual[end_atual/4] ;
            R_atual[end_atual/4] = raio_atual ;
            if (R_ant[end_atual/4] == 0)
                R_ant[end_atual/4] = R_atual[end_atual/4] ;
            raio_medio = (R_atual[end_atual/4] + R_ant[end_atual/4]) / 2 ;
            RQ[npt_atual][end_atual/4] = raio_medio * raio_medio ;
            /*
            /* if (R_ant[end_atual/4] > 0) */
                fprintf(outfile_ptr,"%6.1f ", raio_medio) ; /*
            word = 0;
        }

        if (cont==1)
        {
            y = (RQ[npt_atual][0]-RQ[npt_atual][1])/4060.0;
            z = 586.0105+(RQ[npt_atual][1]-RQ[npt_atual][2]+2030.0*y)/3516.063;
            x = sqrt(RQ[npt_atual][0]-1030225.0-y*y-2030.0*y-z*z);

```

```

        if (R_ant[end_atual/4] > 0)
        {
            fprintf(outfile_ptr,"%6.1f %6.1f %6.1f ", x, y, z);
            fprintf(outfile_ptr,"\n");
        }
    }

    npt_atual = npt_atual - 1;

}

fclose(outfile_ptr);

}

/*                                     */
/* Funcao lenibble                                     */
/*                                     */
/* entrada: endereco do nibble na EPLD, de 0 a 16     */
/* retorno: conteudo do nibble enderecado (8 bits : 0000nnnn) */
/*                                     */

unsigned char lenibble(unsigned char end_nibble)
{
    int nibble, dado, status_port ;
    nibble = 0;

    /* Escr. 1 em LRD (permite leit.) = Escr. 0 em stopen (inibe /INTR) */
    /* Pulso 0-1-0 no gate do lach de entrada da EPLD (BD0) */

    dado = (end_nibble * 16) + 8;          /* +8 para setar bit de LRD, */
                                          /* *16 desloca end. p/ MSB */

    delay(d);
    outportb(0x378, dado);                /* "0" em BD0(/STROBE - EPLD) */
    delay(d);
    outportb(0x378, dado);                /* Repete "0" por garantia */
    delay(d);
    outportb(0x378, dado+1);              /* Nivel "1" em BD0 */
    delay(d);
    outportb(0x378, dado);                /* Nivel "0" em BD0 */
    delay(d);

```

```

status_port = inportb(0x379);          /* le um byte do port de status */
delay(d);

/* Volta endereco para zero */

/*      dado = 8;          */          /* 8 para setar bit de LRD,      */
/*      */          /* *16 desloca end. p/ MSB      */
/*      delay(d);        */          /* "0" em BD0(/STROBE - EPLD) */
/*      outportb(0x378, dado); */
/*      delay(d);        */          /* Repete "0" por garantia     */
/*      outportb(0x378, dado); */
/*      delay(d);        */          /* Nivel "1" em BD0           */
/*      outportb(0x378, dado+1); */
/*      delay(d);        */          /* Nivel "0" em BD0           */
/*      outportb(0x378, dado); */
/*      delay(d);        */          /*

/* Monta o nibble a partir de mascaras para alguns bits de status */

if (!(status_port & 0x80))
    nibble = nibble + 0x08;
if (status_port & 0x20)
    nibble = nibble + 0x04;
if (status_port & 0x10)
    nibble = nibble + 0x02;
if (status_port & 0x08)
    nibble = nibble + 0x01;

return nibble;

}

/*      */          /*
/*      Funcao numera_outfile */
/*      */          /*

int numera_outfile()
{

    char oldname[20] = "" ;
    char newname[20] = "" ;
    char num[5] = "" ;
    char *ext = ".dat" ;
    char dezena, unidade ;

```

```

strcpy(oldname, "xx.dat");
strcpy(newname, nomebase);
dezena = nfile / 10;
unidade = nfile % 10;
num[0] = 48 + dezena;
num[1] = 48 + unidade;
num[2] = '\0';
printf("\n");
/* printf(num); */
printf("\n");
strcat(newname, num);
strcat(newname, ext);
printf("%s => %s\n\n",oldname,newname);

/* Renomeia o arquivo */
if (rename(oldname, newname) != 0)
    perror("rename");
return 0;
}

```

Referências

- [1] J. C. Adamowski, “Transdutores Acústicos e suas Aplicações” , texto apresentado à Universidade Politécnica de São Paulo para fins de Concurso de Livre Docência, Julho de 1995.
- [2] J. C. Adamowski, e outros, “Sistema Computadorizado para Mapeamento de Campo Acústico” , *XIV Congresso Nacional de Ensaaios Não Destrutivos*, Rio de Janeiro, Setembro de 1995.
- [3] J. C. Adamowski, e outros, “Sensores de ultra-som para automação e robótica”, *Revista Máquinas e Metais*, nº 286, pp. 51-55, Novembro de 1989.
- [4] B. Agranat, e outros, “*Fundamentos de la Fisica y la Tecnica de los Ultrasonidos*”, Editora Mir, Moscou, 1990.
- [5] J. A. Anderson, “*An Introduction to Neural Networks*”, MIT Press, 1995.
- [6] S. Ando, “An Autonomous Three-Dimensional Vision Sensor with Ears”, *IEICE Transactions on Inf. & Syst.*, vol. E78-D, nº 12, pp. 1621-1629, December 1995.
- [7] B. Barshan, “A Bat-Like Sonar System for Obstacle Localization”, *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 22, nº 4, pp. 636-647, July/August 1992.
- [8] S. S. Blackman, *Multiple Target Tracking with Radar Applications* , Artech House, 1986.
- [9] L. Camp, “*Underwater Acoustics*”, Wiley-Interscience, 1970.
- [10] C. Canali, et al, “A Temperature Compensated Ultrasonic Sensor Operating in Air for Distance and Proximity Measurements” , *IEEE Transactions on Industrial Eletronics*, vol. IE-29, nº 4, November 1982.
- [11] J. W. Choi, Y. H. Kim, “Spherical Beam-forming and Music Methods for the Estimation of Location and Strength of Spherical Sound Sources”, *Mechanical Systems and Signal Processing*, Academic Press, May 1995.
- [12] J. H. Clements III, J. T. Lo, “Recursive Direction Finding in the Presence of Sensor Array Uncertainties”, *Proceedings of the IEEE*, April 1993.
- [13] F. Damiani, J. Manzolli, G. Mendes, “Controle Paramétrico Midi usando Interface Gestual Ultrassônica”, V Simpósio Brasileiro em Computação e Música, XVII Congresso da SBC, UFMG, Belo Horizonte, pg. 55-60, 1998.
- [14] M. S. Datum, F. Palmiere, A. Moiseff, “An Artificial Neural Network for Sound Localization Using Binaural Cues”, *J. Acoust. Soc. Am.*, vol. 100, nº 1, pp. 372-383, July 1996.

- [15] D. J. DeFatta, J. G. Lucas, W. S. Hodgkiss, “*Digital Signal Processing : A System Design Approach*” , Wiley, 1988.
- [16] R. M. Eisberg, L. S. Lerner, “*Física – Fundamentos e Aplicações*”, McGraw-Hill, 1983.
- [17] J. A. Fabro, “Grupos Neurais e Sistemas Nebulosos: Aplicação à Navegação Autônoma” , *Tese de Mestrado*, Universidade Estadual de Campinas, 1996.
- [18] F. Figueroa, E. Barbieri, ”An Ultrasonic Ranging System for Structural Vibration Measurements” , *IEEE Transactions on Instrumentation and Measurement*, vol. 40, nº 4, pp. 746-769, August 1991.
- [19] J. J. Fuchs, “Shape calibration for a nominally linear equispaced array”, *Proceedings of the IEEE*, abril de 1993.**
- [20] K. Fukushima, S. Miyake, T. Ito, “Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition” , *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13, pp. 826-834, 1983.
- [21] P. M. Gammell, “Improved Ultrasonic Detection Using the Analytic Signal Magnitude” , *Ultrasonics*, March 1981.
- [22] P. M. Gammell, “Analogue Implementation of Analytic Signal Processing for Pulse-Echo Systems” , *Ultrasonics*, November 1981.
- [23] G. C. Gaunaud, H. C. Strifors, “Signal Analysis by means of Time-Frequency (wigner-Type) Distributions – Applications to Sonar and Radar Echoes” , *Proceedings of the IEEE*, vol.84, nº 9, September 1996.
- [24] R. P. Gooch, J. J. Shynk, “Wide-band Adaptive Array Processing Using Pole-zero Digital Filters” , *IEEE Transactions on Antennas and Propagation*, March 1986.
- [25] M. T. Hagan, “Training Feedforward Networks with the Marquardt Algorithm” , *IEEE Transactions on Neural Networks*, vol. 5, nº 6, November 1994.
- [26] M. H. Hassoun, *Fundamentals of Artificial Neural Networks* , The MIT Press, 1995.
- [27] S. Haykin, A. Steinhardt, *Adaptive Radar Detection and Estimation*, JohnWiley & Sons, 1992.
- [28] S. Haykin, “*Neural Networks - A Comprehensive Foundation*”, Prentice Hall, 1994.
- [29] S. Haykin, “Neural Networks Expands Signal Processing Horizons” , *IEEE Signal Processing Magazine*, March 1996.
- [30] S. Haykin, “*Adaptive Filter Theory*”, Prentice Hall, 1996.

- [31] J. Hertz, A. Krogh, R. G. Palmer, “*Introduction to the Theory of Neural Computation*”, Addison Wesley, 1991.
- [32] J. Hollan, et al, “An Introduction to HITS: Human Interface Tool Suite” , *Intelligent User Interfaces*, J. W. Sullivan, S. W. Tyler, editors, ACM Press, 1991.
- [33] H. Demuth, M. Beale, “*Neural Network Toolbox*”, The Math Works Inc., 1994
- [34] D. R. Hush, B. G. Horne, “Progress in Supervised Neural Networks”, *IEEE Signal Processing Magazine*, Jan 1993.
- [35] L. Kleeman, R. Kuc, “Mobile Robot Sonar for Target Localization and Classification”, *The International Journal of Robotics Research*, vol. 14, nº 4, pp. 295-318, August 1995.
- [36] W. C. Knight, R. G. Pridham, S. M. Kay, “Digital Signal Processing fo Sonar” , *Proceedings of the IEEE*, vol. 69, nº 11, November 1981.
- [37] B. Kosko, editor, “*Neural Networks for Signal Processing*”, Prentice Hall, 1992.
- [38] H. Krim, M. Viberg, “Two Decades of Array Signal Processing Research”, *IEEE Signal Processing Magazine*, July 1996.
- [39] S. Kuroyanagi, A. Iwata, “Auditory Pulse Neural Network Model to Extract the Inter-Aural Time and Level Difference for Sound Localization”, *IEICE Transactions on Inf. & Syst.*, vol E77-D, nº 4, pp. 466-474, April 1994.
- [40] G. Kurtenback, E. A. Hulteen, “Gestures in Human-Computer Communication”, *The Art of Human-Computer Interface*, B. Laurel, editor. Addison-Wesley, 1993.
- [41] L. C. Lynnworth, *Ultrasonic Measurements for Process Control: Theory, Techniques, Applications*, Academic Press, 1989.
- [42] A. Marsal, S. Marcos, “A Criterion to Detect Sensor Array Distortion”, *Proceedings of the IEEE*, Abril de 1993.
- [43] P. Mattila, M. Luukkala, “FET Pulse Generator for Ultrasonic Pulse Echo Applications”, *Ultrasonics*, 1981.
- [44] G. Mendes, F. Damiani, “Interface Gestual Ultrassônica”, *IV Workshop Iberchip*, Mar del Plata, Argentina, 1998.
- [45] A. Mulder, “Human Movement Tracking Technology” , Technical Report 94-1, *Hand Centered Studies of Human Movement Project*, School of Kinesiology, Simon Fraser University, 1994.

- [46] A. Mulder, “Virtual Musical Instruments: Accessing the Sound Synthesis Universe as a Performer” , School of Kinesiology, Simon Fraser University, 1994.
- [47] P. O. Navaux, F. S. Osório, R. V. Dorneles, “Processamento de Sinais Através de Redes Neurais”, *Workshop sobre Computação de Alto Desempenho para Processamento de Sinais*, 1993.
- [48] L. X. Nepomuceno, “*Tecnologia Ultra-Sônica*”, Edit. Edgar Blücher, 1980.
- [49] H. Nonaka, T. Da-te, “Ultrasonic Position Measurement and Its Application to Human Interface”, *IEEE Transactions on Instrumentation and Measurement*, vol. 44, no. 3, pp. 771-774, June 1995.
- [50] T. Oenning, “Experiments in Beamforming and CDMA Using an Ultrasound Sensor Array”, *Senior Honors Project*, University of Minnesota, May 1995.
- [51] G. Papadopoulos, et al, “Implementation of an Intelligent Instrument for Passive Recognition and Two-Dimensional Location Estimation of Acoustical Targets”, *IEEE Transactions on Instrumentation and Measurement*, vol. 41, nº 6, pp. 885-890, December 1992.
- [52] M. Parrilla, J. J. Anaya, and C. Fritsch, “Digital Signal Processing Techniques for High Accuracy Ultrasonic Range Measurements”, *IEEE Transactions on Instrumentation and Measurement*, vol. 40, nº 4 , pp. 759-763, August 1991.
- [53] P. A. S. Pereira, “Sensor Ultrassônico Autônomo de Exploração de Entornos Orientado a Problemas de Navegação” , *Trabajo Fin de Maestría*, Universidad Politécnica de Madrid, 1993.
- [54] S. U. Pillai, “Array Signal Processing”, Springer-Verlag, 1989.
- [55] C. Pohlenz, “Power Electronic Devices for Sonar Systems”, Groupe d’Etude et de Recherche en Détection Sous-Marine (GERDSM), Toulon, France , >1986
- [56] A. M. Sabatini, “A Digital Signal-Processing Technique for Compensating Ultrasonic Sensors”, *IEEE Transactions on Instrumentation and Measurement*, vol. 44, nº 4, pp. 869-874, August 1995.
- [57] P. K. Simpson, “Neural Networks for Sonar Signal Processing” , *Handbook of Neural Computing Applications*, >1990.
- [58] S. Talwar, A. Paulraj, G. H. Golub, “A Robust Numerical Approach for Array Calibration”, *Proceedings of the IEEE*, April 1993.
- [59] F. J. Von Zuben, “Redes Neurais Aplicadas ao Controle de Máquinas de Indução”, *Tese de Mestrado*, Universidade Estadual de Campinas, 1993.

- [60] F. J. Von Zuben, “Modelos Paramétricos e Não-Paramétricos de Redes Neurais Artificiais e Aplicações”, *Tese de Doutorado*, Universidade Estadual de Campinas, 1996.
- [61] D. Webster, “A Pulsed Ultrasonic Distance Measurement System based upon Phase Digitizing” , *IEEE Transactions on Instrumentation and Measurement*, vol. 43, nº 4, pp. 578-582, August 1994.
- [62] D. Weimer, S. K. Ganapathy, “Interaction Techniques Using Hand Tracking and Speech Recognition” , *Multimedia Interface Design*, M. Blattner, R. B. Dannenberg, editors, ACM Press, 1992.
- [63] A. Wexelblat, “Natural Gesture in Virtual Enviroments” , *ACM Virtual Reality Systems and Technology(VRST'94) Proceedings*, ACM Press,1994.
- [64] J. S. K. Woo, “A Short History of the Developments of Ultrasound in Obstetrics and Gynecology”, <http://home.hkstar.com/~joewoo/history2.html>, 1998.
- [65] P. Wu, R. Kazys, T. Stepinski, “Analysis of the Numerically Implemented Angular Spectrum Aproach based on the Evaluation of 2-D Acoustic Fields”, Parts 1 and 2, *Journal of Acoustic Society of America*, March 1996.
- [66] M. Yang, et al, “A Multifrequency AM-Based Ultrasonic System for Accuracy Distance Measurement”, *IEEE Transactions on Instrumentation and Measurement*, vol. 43, nº 6, pp. 861-866, December 1994.