



ANÁLISE DOS SINAIS DE VOZ NOS DOMÍNIOS DO TEMPO E DA FREQUÊNCIA

Roberto Miscow Filho

Em 1822, o matemático Joseph Fourier publicava a monografia "Teoria Analítica do Calor" na qual apresentava o estudo sistemático do desenvolvimento de funções em séries de senos e cossenos. Tal estudo hoje em dia, muito mais desenvolvidos, constitui a chamada "análise de Fourier" e sua aplicação abrange a interpretação de sinais periódicos no tempo, sinais de duração limitada e, até mesmo, sinais aleatórios (com respeito aos sinais aleatórios, são fundamentais os estudos de Norbert Wiener, o pai da cibernética).

Com o rápido desenvolvimento das modernas técnicas eletrônicas digitais (tornadas mais facilmente realizáveis por meio dos transistores e dos circuitos integrados), cresceu paralelamente a importância dos sinais amostrados, isto é, dos sinais em tempo "discreto"

(em oposição a tempo contínuo). Como consequência do mesmo fenômeno tecnológico (crescimento das técnicas digitais), expandiu-se a análise de Fourier aos ditos sinais amostrados, aos sinais que ocorrem em tempos discretos.

O núcleo conceitual da análise de Fourier, isto é, seu aspecto teórico mais importante, constituiu-se na seguinte propriedade, tão simples quanto elegante:

- "existe uma relação bi-unívoca entre um sinal temporal (seja o tempo contínuo ou discreto) e seu espectro, isto é, o conteúdo de frequência componentes do sinal; em outras palavras: a cada sinal função do tempo, t , corresponde um único espectro, função da frequência, f ".

Esse relacionamento bi-unívoco permite ao engenheiro, ao cientis-

ta, ao pesquisador, interpretar um sinal através do espectro respectivo e, através dessa interpretação, chegar às raízes do fenômeno físico que gerou aquele sinal.

O presente artigo pretende apresentar uma das mais interessantes aplicações da análise de Fourier qual seja: o estudo dos sinais oriundos da voz humana.

Alguns Exemplos

— O que vai ser mostrado a seguir é apenas uma amostra muito pequena de resultados obtidos no Instituto Militar de Engenharia, em laboratório de sua Seção de Engenharia Elétrica. Utilizou-se, nos trabalhos feitos, o equipamento analisador espectral tipo 2033 da BRUEL & KJAER, acoplado a um registrador X-Y, tipo 2308, da mesma indústria.

O referido analisador usa a técnica digital mais moderna, em particular a chamada "Transformada Rápida de Fourier" (Fast Fourier Transform-FFT); sua extrema flexibilidade permite a representação visual do sinal no domínio do tempo ou de seu respectivo espectro, este último em uma de várias possíveis escalas de freqüência, incluindo a possibilidade de ampliação em uma janela espectral (ampliação essa denominada "zoom").

Apesar de sua grande flexibilidade, o analisador 2033 permite o registro de um sinal cuja duração seja no máximo de oito décimos de segundo (0,8 s). Para o leitor não acostumado com medições desse tipo, esclarecemos que uma

frase coloquial, por exemplo: "Papai, quem fala aqui é o Pedro!" ou "Espero você no sábado que vem!", dura em média 2 a 3 segundos. Considerando, pois, a limitação temporal do analisador, tivemos que operar sobre sílabas e palavras isoladas. Para sinais de maior duração, brevemente estaremos utilizando no IME um terminal para aquisição de dados constituído por um microcomputador e um conversor analógico/digital, terminal esse projetado na própria Seção de Engenharia Elétrica do Instituto.

As palavras utilizadas nos testes que forneceram o assunto do presente artigo foram as seguintes:

CADETE
BÉLICO
MUNIÇÃO
TIRO

A primeira dessas palavras, quando enunciada ao microfone em entonação convencional (isto é, sem afetar surpresa ou dúvida) gerou um sinal cuja representação gráfica no tempo aparece reproduzida na Figura 1. A observação da figura permite ver nitidamente o trecho correspondente à sílaba [ca], o trecho correspondente à sílaba [dê] e, finalmente o trecho da sílaba [te]. A escala horizontal varia de zero a oito décimos de segundo (0 a 0,8 s).

Pode-se notar que a sílaba átona [te] contém bem menos energia; precedendo a sílaba [dê], percebe-se um trecho de pequena amplitude no qual o sinal é quase periódico. Esse trecho corresponde à oscilação da consoante vozeada [d]

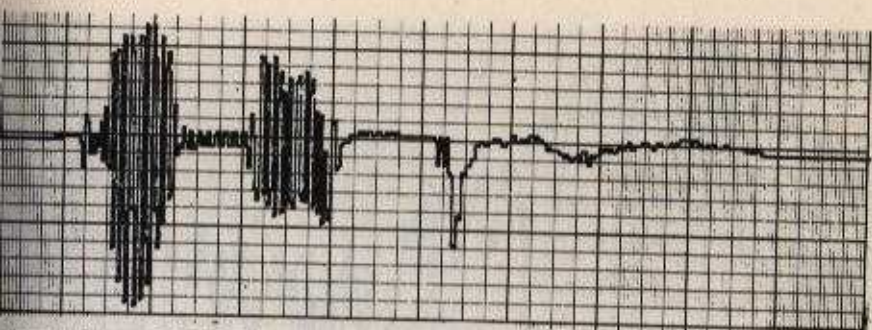


Figura 1: Palavra CADETE (escala: 0 a 0,8 s).

(consoante vozeada é aquela que depende das cordas vocais).

Aliás, é preciso notar que as consoantes (vozeadas e não vozeadas) geram sinais cuja forma depende do fonema vocálico associado; esse fato pode ser verificado pelo próprio leitor prestando atenção à configuração tomada pela boca ao se preparar para enunciar um som: a configuração de um [ca], por exemplo, é diferente da

correspondente a um [dê], a um [di] etc.

O analisador usado em nossas experiências permite uma observação mais detalhada dos sinais no domínio do tempo porquanto um dos recursos do aparelho é a possibilidade de expandir a escala do tempo, isto é, passar, p.ex., de uma escala em segundos para uma escala em milisegundos. Desse modo, pode-se por exemplo, ob-

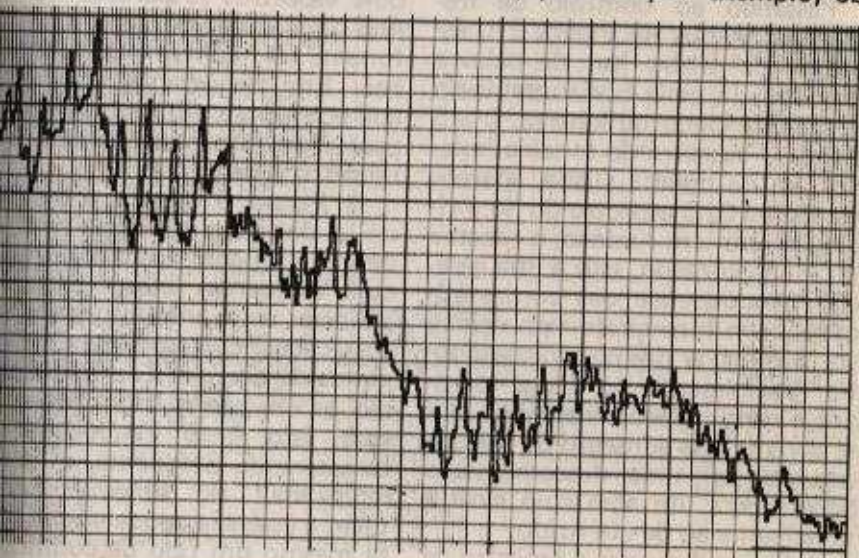


Figura 2: Espectro da sílaba [ca] (escala: 0 a 6 KHz).

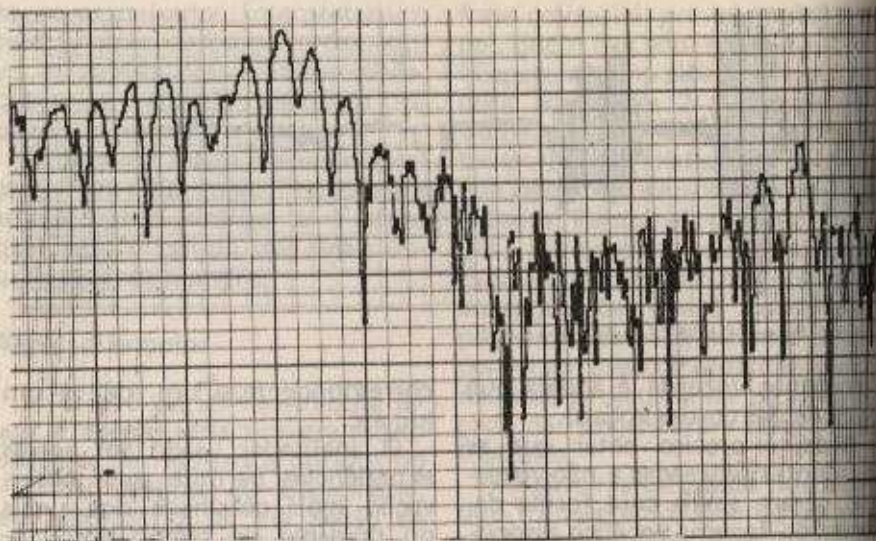


Figura 3: Espectro de [k] da sílaba [ka] (escala: 0 a 5 KHz).

servar isoladamente o efeito de um fonema consonantal: o [k] da sílaba [ka], o [t] da sílaba [te], o [m] da sílaba [mu] etc.

Passando ao domínio da frequência, vemos na Figura 2 o espectro da sílaba [ka] da palavra "cadete" (a escala horizontal varia de zero a cinco quilohertz, 0 a 5 KHz). O espectro do fonema isolado [k], da mesma sílaba, aparece na Figura 3. O espectro do [k] é típico de um som ruidoso; o fonema [k] é não-vozeado, isto é, sua geração não depende das cordas vocais. É um fonema do tipo oclusivo, isto é, o ar dos pulmões encontra, inicialmente, os lábios fechados e depois, repentinamente, eles se abrem. Esse aumento de pressão seguido por uma descompressão brusca gera um som cujo sinal é semelhante ao de um ruído.

A Figura 4 mostra o espectro da sílaba [de] da palavra "cadete" e a

Figura 5 mostra o espectro da sílaba [te], da mesma palavra. Se o leitor comparar as duas figuras, notará que nas baixas frequências os dois espectros se parecem muito; esse fato é devido à presença do fonema [ê], comum às duas sílabas. Em frequências mais altas, aparecem mais diferenças causadas pelos fonemas consonantais diferentes: o [d] e o [t].

A Figura 6 mostra o espectro do [a] da sílaba [ka] da palavra "cadete". Na figura estão assinaladas, com pequenas setas, certas frequências cujas amplitudes são bem destacadas, correspondendo a máximos de energia. Tais frequências são os chamados "formantes" do som, isto é, as frequências que caracterizam o fonema, a "marca registrada" do fonema. Constituem a essência daquele som. Existe mesmo um tipo de codificação do sinal de voz que aproveita pr

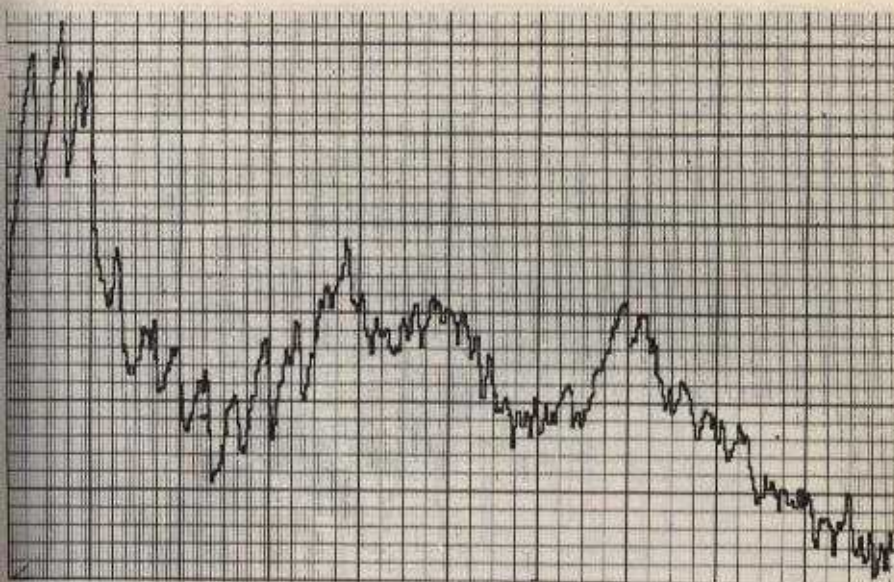


Figura 4: Espectro da sílaba |dê| (escala: 0 a 5 KHz).

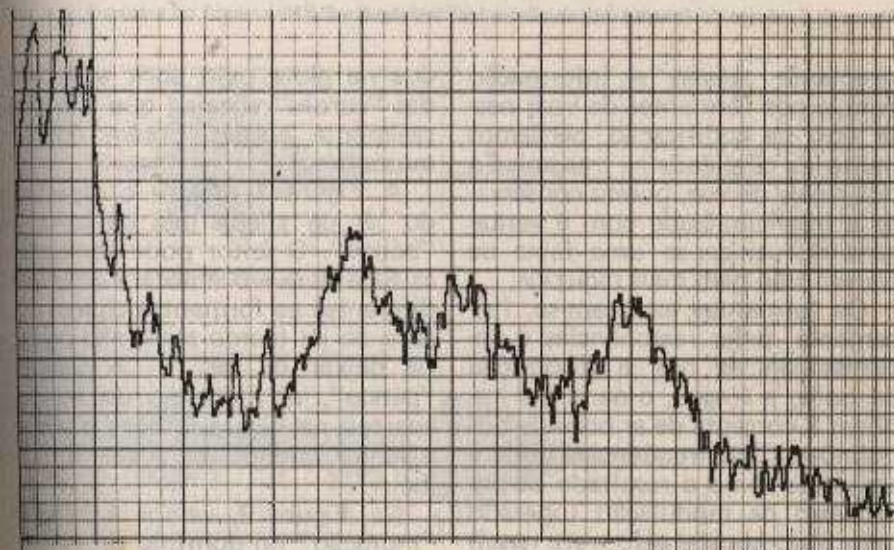


Figura 5: Espectro da sílaba |tê| (escala: 0 a 5 KHz).

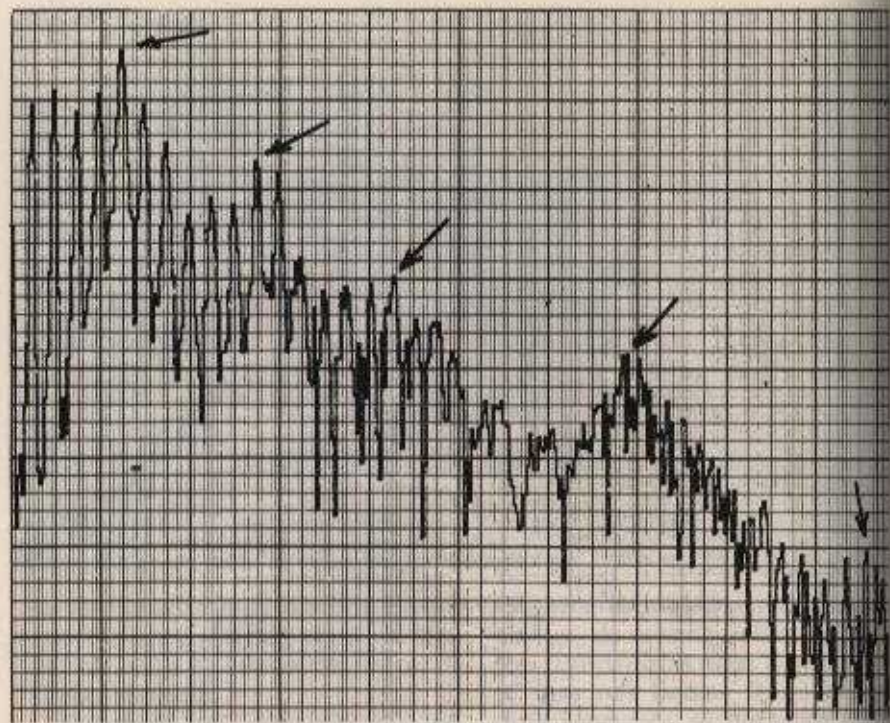


Figura 6: Espectro do fonema [a], da sílaba [ca] (escala: 0 a 5 KHz).

ticamente apenas a informação conida nos formantes da voz; nesse tipo de codificação, perdemos muito da qualidade na transmissão dos sinais de voz, porém conserva-se a inteligibilidade com a vantagem de utilizar-se uma faixa de transmissão bem mais reduzida.

Os formantes explicam-se, fisicamente, pelas ressonâncias e antirressonâncias que a disposição da boca, a língua, o nariz etc impõem à corrente de ar oriunda dos pulmões. Tais frequências podem ser vistas como o resultado de um processo de modulação (operado pelo trato bucal e pelo trato nasal) nos pulsos periódicos de ar gera-

dos na glote pelo abrir e fechar das "cordas" vocais (que não são realmente cordas, porém dobras musculares).

A Figura 7 mostra o espectro do [é] da sílaba [be] da palavra "bélico". O leitor pode comparar as Figuras 5 e 6 e notar como são diferentes os formantes típicos do [a] e do [é] (como aliás, não poderia deixar de ser...). A palavra "bélico", no domínio do tempo aparece na Figura 8; note o leitor a maior energia da sílaba tônica [bé].

Na Figura 9, vemos a palavra "munição" no domínio do tempo (escala de zero a oito décimos de

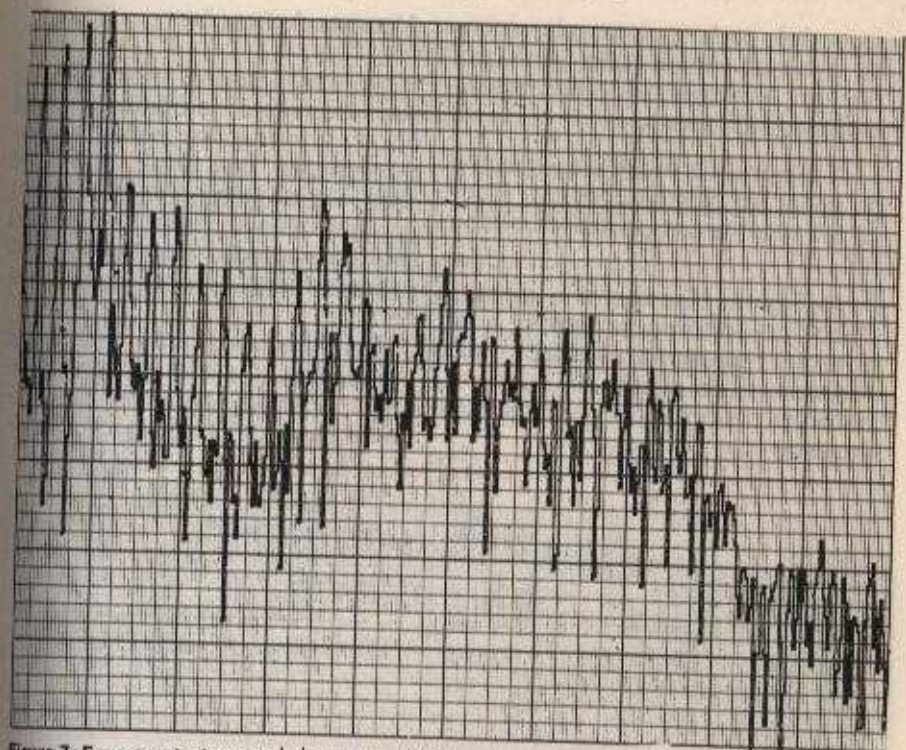


Figura 7: Espectro do fonema [é], da sílaba [bé] (escala: 0 a 5 KHz).

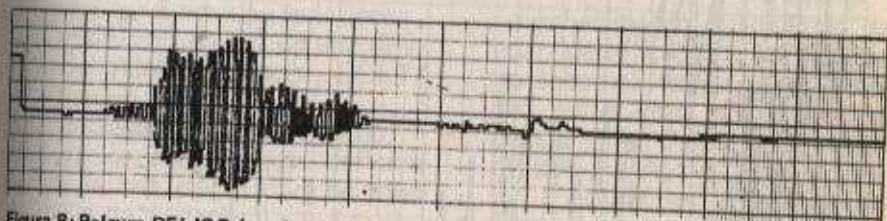


Figura 8: Palavra BELICO (escala: 0 a 0,8 s).

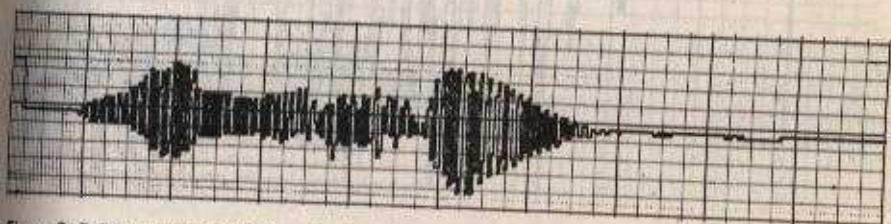


Figura 9: Palavra MUNIÇÃO (escala: 0 a 0,8 s).

segundo); pode-se notar, claramente, a sílaba tônica "cão", iniciada por um som tipo ruído correspondente ao sibilante [ç].

A Figura 10 mostra o espectro da vogal [u] da sílaba [mu]; a Figura 11 mostra o espectro da sílaba [ni].

A Figura 12 mostra a palavra "tiro" no domínio do tempo; a Figura 13 representa o espectro do fonema [t] da sílaba [ti]. Notar a natureza ruidosa do fonema não vozeado [t].

Algumas Aplicações

No texto supra, mostramos alguns exemplos de como é possível obter informações sobre o comportamento dos sinais de voz no domínio do tempo e no domínio da frequência. Tais informações são aproveitadas intensamente no processamento digital daqueles sinais.

Por sua vez, o processamento digital dos sinais de voz tem diversas aplicações, podendo ser desdobradas as seguintes:

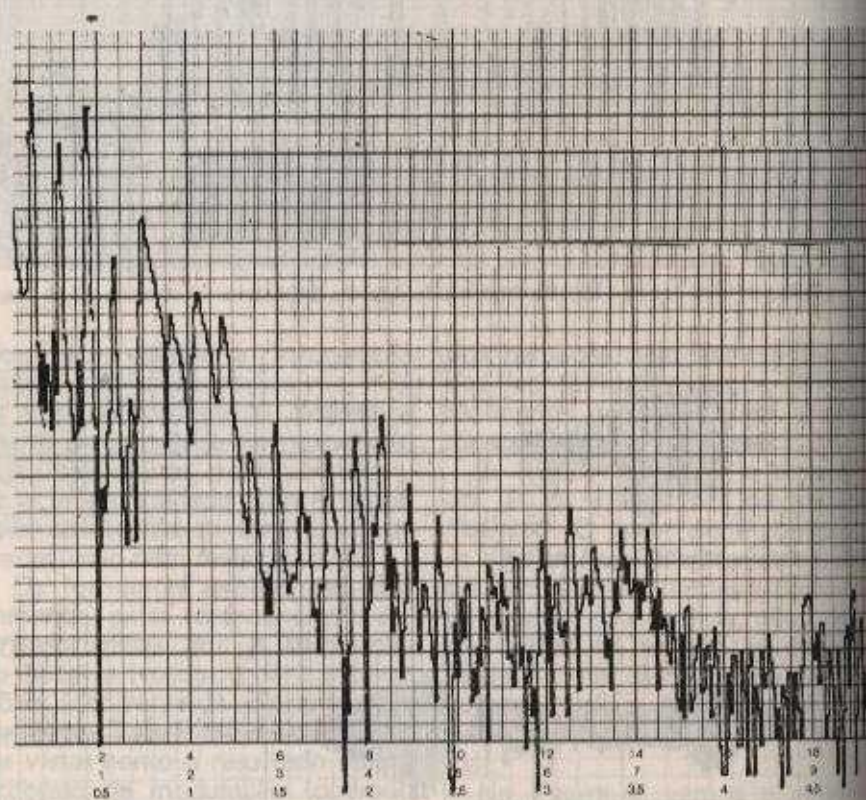


Figura 10: Espectro do fonema [u] da sílaba [mu] (escala: 0 a 5 KHz).

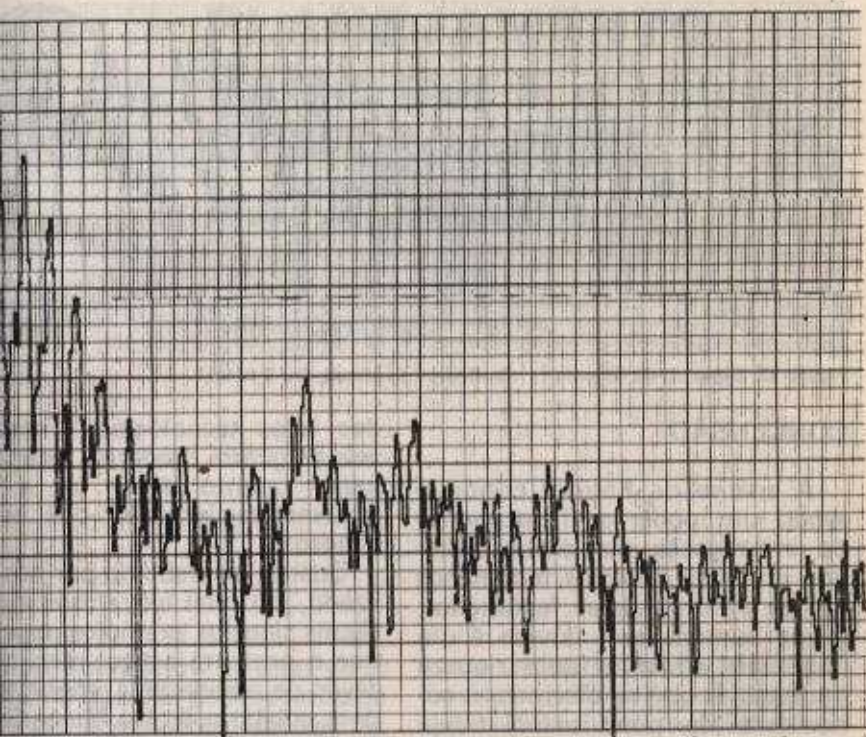


Figura 11: Espectro da sílaba 'ni' (escala: 0 a 5 KHz).

pesquisas e estudos fonéticos;
transmissão digital da voz (Sistema PCM diferencial, DELTA adaptivo etc);
armazenagem de informação;
recursos para pessoas deficientes da fala, por motivo de doença ou de acidente;
comunicação direta homem-máquina;

— reconhecimento automático da voz (identificação e verificação do locutor) (sobre essa aplicação publicamos um artigo na revista "A Defesa Nacional" nº 696-Jul/Ago 81, sob o título: "Aplicações Militares do Reconhecimento Automático da Voz").

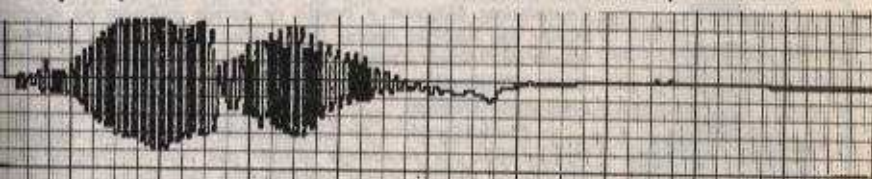


Figura 12: Palavra TIRO (escala: 0 a 0,8 s).

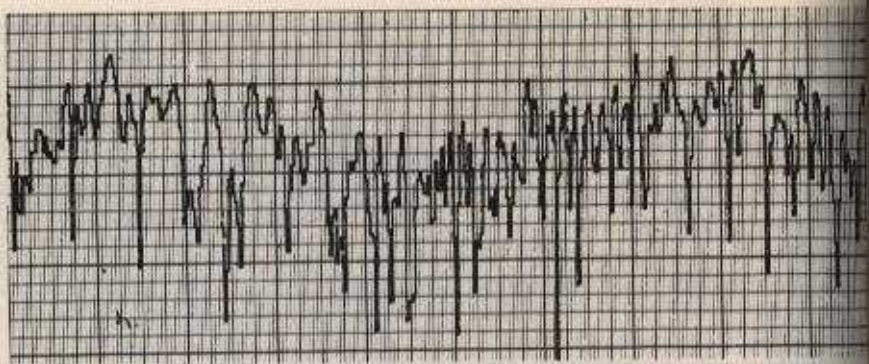


Figura 13: Espectro do fonema [t] da sílaba [ti] (escala: 0 a 5 KHz)

CONCLUSÃO

A análise de Fourier continua sendo um poderoso instrumento de investigação dos sinais oriundos dos mais diversos fenômenos físicos e, entre esses fenômenos, a voz humana sem dúvida é um dos mais atraentes pela riqueza de detalhes que se oferecem à curiosidade do pesquisador, esteja ele voltado para alguma aplicação prática ou interessado apenas na simples descoberta.

Acreditamos que aquele capítulo emocionante da matemática superior precisa ser do conhecimento do nosso corpo de oficiais, sejam eles engenheiros ou combatentes, por quanto a compreensão

dos fatos tecnológicos, inseparáveis da guerra moderna, será tanto maior quanto maior for a base matemática-científica dos nossos quadros.

Quanto ao trabalho por nós realizado, do qual extraímos apenas alguns singelos exemplos para ilustrar o presente artigo, resta-nos o dever de agradecer a importantíssima colaboração do Major General Magela Pinheiro Gomes, nosso aluno de Pós-Graduação no IME, bem como agradecer, também, a cooperação da TELERJ na pessoa do Dr. D'Alpoim Beda dos Reis por nos haver permitido o uso das câmaras acústicas daquela empresa, nas quais pudemos fazer gravações sem a interferência de perturbadores ruídos ambientais.



O Cel. ROBERTO MISCOW FILHO pertence ao Quadro de Engenheiros Militares e possui os cursos militares da Academia Militar das Agulhas Negras (Infanteria), da Escola de Aperfeiçoamento de Oficiais (Comunicações) e do Instituto Militar de Engenharia (Engenharia de Comunicações e Mestre em Ciências de Engenharia Elétrica).

Prestou serviços no 13º Batalhão de Caçadores, Joinville SC (1953-1956), na Academia Militar das Agulhas Negras (1957-1958) e no Serviço Rádio do Ministério do Exército (1963-1968). Atualmente é o Chefe da Divisão de Ensino e Pesquisa do Instituto Militar de Engenharia (IME).