SOUND IMAGE RECOGNITION BY HOLOGRAPHIC MEANS

V.R. IMAKOV, V.N. SOBOLEV

All-Union By Correspondence Electrotechnical Institute of Communications, Moscow, USSR 123855

ABSTRACT

Optical methods of sound image recognition are discussed as an alternate to recognition systems with von Neumann's architecture. The main design principles and algorithms are described.

INTRODUCTION

Most sound image recognition systems are based on computer systems (CS) with a von Neumann architecture (with a sequential instruction stream). As is known such CS are not equipped with functions (including input-output) required to process non-numeric data, such as speech, graphic images, etc. Due to the difficulties of real-time parallel processing of acoustic signals, it appears expedient to shift to customized optical computers (COC) for solving sound image recognition problems. Such COC may use both coherent and non-coherent light emissions, or their combination. Correlation type COC are the most widely used due to the simplicity and efficiency of complex signal transformations, such as convolution, correlation, Fourier transforms, Hankel transforms, multiplication of matrices, etc. In contrast to traditional digital CS in which the elementary operation is a comparison by mod 2, in COC of the correlation type an elementary operation is a complex functional or integral transformation with an execution time determined only by the time of light travel through the optical media and assemblies, which can be some 10 ns to 10 ps. Another feature of COC is their ability to perform multiparallel signal processing. Optical computers are equivalent to CS with 10^6 to 10^{12} inputs. The number of output channels can range from 1 to kn, where n is the number of inputs and k = 1, 2,3... Another feature of COC is the ease and simplicity of processing multidimensional objects. COC which should simultaneously, in fractions of a microsecond add and divide hundreds of millions of numbers or multidimensional matrices can be designed without undue engineering problems, while such speeds in traditional digital CS are unattainable, especially if the simultaneous processing of great data bulks is taken into account. Many researches tend to treat acoustic signals as a unidimensional $f_t(x)$ one, rather than three-dimensional $f_t(x,y,z)$ signals, this being hardly always justified. Holography is an ideal means of mathematical simulation of three-dimensional objects, and holographic methods provide an exhaustive description of acoustic signals at all stages of its processing.

IMPLEMENTATION

Optical computers can be designed as analog, digital, or hybrid devices and may include various electronic and mechanical assemblies and units. COC functioning is based on the principle of generalised image delineation. The acoustic signal is fed to optical channel mostly via the so-called spatial-time light modulators (STLM) in the form of a two- or three-dimensional matrix consisting of several hundred or thousand cells controlled by the acoustic signal or its electric equivalent. The acoustic or electric signal causes a charge image to be formed on the modulator surface and this in turn modulates the light beam. STLMs may be operated both in the light transmission or light reflection modes. One of the STLM modifications is the controlled liquid crystal matrix (with acoustic, electric, or light control) with modulation frequencies up to about 60 kHz which is usually adequate for acoustic signal processing. The most advanced acoustic light modulators (of the Phototitus type) are based on CRTs [1,2], with a special crystal serving as the target inside the CRT and two electron guns for information recording and erasure, respectively. The charge image pattern on the crystal surface is formed by a controllable electron beam. During information readout the passing coherent light is phase and amplitude modulated. Real-time operation is provided by a second electron gun with a wide beam to remove the surface charge. As demonstrated [2], noncoherent optical processing is essentially reduced to linear operations with the image. In the classical non-coherent optical processor [3] the correlated output signal appears on a background of a constant bias. In the past, applications of such systems have been hampered by the low output signal-to-noise ratio and the difficulties of handling complex data. In the non-coherent optical speech processing system under study a higher signal-to-noise ratio is obtained and the constant bias is eliminated by modulating and demodulating the carrier. This makes it feasible to preprocess complex data to a form suitable to be input to the main coherent processor; this is accomplished with the aid of an obscure aperture of special shape. Consider two-dimensional functions: the recognized acoustic pattern $f(x,y)$ and the reference pattern $g(x,y)$ which are to be compared by closeness. In the general case, they can be complex quantities. Using their optical image, coded transparencies with transmission intensities f_c and g_c are generated:

$$f_c = 0.5|f(x_1,y_1)|\{1+\cos[2\pi\nu_c x_1 + \arg f(x_1,y_1)]\} \quad (1)$$

$$g_c = 0.5|g(x_1,y_1)|\{1+\cos[2\pi\nu_c x_1 + \arg g(x_1,y_1)]\} \quad (2)$$

where ν_c is the carrier frequncy used in the coding operation. Functions f_c and g_c are realized as intensities caused by biasing the cosine carrier. At $|f| \le 1$ and $|g| \le 1$, we have $0 \le |f_c| \le 1$ and $0 \le |g_c| \le = 1$. This means that processing the coded transparencies is equivalent to processing the initial functions. Correlation between f_c and g_c is provided by the base non-coherent processor (Fig. 1). Fresnel holograms for the plane of obscure P' were generated, with transmittance functions $g_c(x_1,y_1)$ in the input plane P_1 corresponding to various phonems and their combinations (dyads). The transparency modulated by f_c was positioned in the P_1 plane and thus the light intensity in the output plane P_1 was $f_c \circledast g_c$:

$$I_2 = f_c \circledast g_c = 0.25\ |f| \circledast |g| + 0.25\ |f \circledast g| \cos[2\pi\nu_c x_2 + \arg(f \circledast g)] + 0.25\ |f| \circledast |g| |\cos[2\pi\nu_c x_2 + \arg(g)]| + 0.25\ |g| \circledast |f| \quad (3)$$

If ν_c is sufficiently large, the signal spectra of main frequency band with a modulated carrier in Eqs. (1) and (2) will not overlap in the frequency domain. Since correlation is equivalent to multiplication in the frequency domain, the last two terms in Eq. (3) will be zero and the pattern in plane P_2 will be reduced to:

$$I_2 = f_c \circledast g_c = 0.25\ |f| \circledast |g| + 0.25\ |f \circledast g| \cos[2\pi \nu_c x_2 + \arg(f \circledast g)] \qquad (4)$$

To obtain the desired complex function $f \circledast g$ from the distribution in the P_2 plane the pattern in this plane was scanned by a raster in the x_2 direction, with the spatial carrier ν_c being transformed into a time carrier $S\nu_c$ (S is the scanning speed). Passing the video signal through a band-pass filter removes the first term of Eq.(3) and the second term then depicts the absolute value and phase of the $f \circledast g$ signal. In the transformation device used masks for the DC component and first, third, fifth and seventh derivatives of the spatial-temporal acoustic signal were provided, with the even derivatives zeroed out by an appropriately selected obscure function. Differentiation and averaging were holographic.The masks were programmed to provide a pseudo-formant representation of the speech signal, this ensuring an adequate invariance relative to different dictors. Pseudo-formants are more descriptive than formants, least of all prone to change, and are relatively easy to separate [4]. Non-coher-

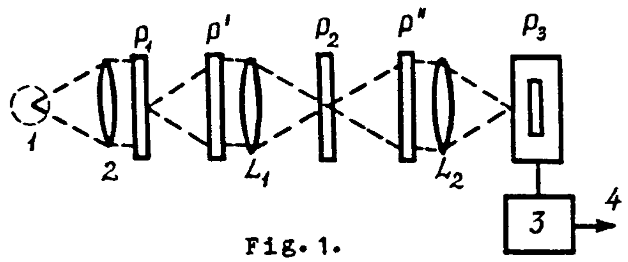

Fig. 1.

Basic non-coherent optical correlator
1 - source; 2 - condenser lens; 3 - decoder; 4 - output signal; P', P" - obscures; P_1, P_2 - mask-transparencies; L_1, L_2 - lenses; P_3 - integral matrix

ent optical speech processing is limited to linear transforms only. Nonlinear transforms of acoustic signals are readily produced by coherent optics techniques, using the "Kristal" facility with a "Prototitus" modulator. Recognition was performed using the multirange delineation and modified image disfocus methods [2,5].

PRINCIPLES OF ALGORITHM CONSTRUCTION

Delineation of "visible speech" patterns by means of a controlled photoelectrooptical liquid-crystal matrix is based on the photosensitive surface being exposed both to a focused image and defocused image, the former providing a pulse response in the form of a delta function and the latter - in the form

$$1/(R_o^2 \mathrm{ciRc}(\sqrt{x^2 + y^2}/R_o)),$$

where R_o is the defocusing factor. The contour is determined by the difference between these images which is generated during readout. Such processing is analoguous to photography with an "unsharp mask" [3].Generalizing Casasent's transform [2,6] by introducing normalization to time and combining geometric transformations with integrated optical processing provides addressing a considerably wider class of phonem speech decoding problems, in particular by including "visible speech" image recognition when the pattern differs from the reference one in scale, positioning, orientation and time dependence. A multigraph is generated in the COC memory as result of holographic speech signal processing, this multigraph containing various interpretations of the recognized words, syllables and phonems. Studies show the optimal recognition algorithm to correspond to the minimal evaluation by Kolmogorov's intricacy criterion. Some relations, describing associative signs are outlined from the versatile relations class. The effects of actually implemented algorithms on the

image being recognized is limited to the screening operator which is in the form of a special mask and which is equivalent in effect to convolution of an associatived sign matrix with a versatile relations matrix. In the intelligent system thus created particular calculus of natural deductions is widely employed. Digital holography was used to design the optimal filter, the initial data being produced by passing the visible speech images through special masks, such as chess field, concentric alternating dark and light bands, moire grid, etc. Computer processing of these prefiltered images produced a program of grid plotting for a precision plotter, with a photo image of this grid reduced by 70X used as an optimal matched filter. The same program was used to control the electron beam path during readout of the recognized visual speech image. Beam deflection was corrected by means of a special associative mask which served as a multiversion prompter. The most probable beam paths were run first with less probable paths following. The artificial intelligence system made wide use of contiguity and hint relations. As compared to frame artificial intelligence systems, this system features the advantages of associative links and a considerably higher version search rate for speech pattern recognition.

FURTHER DEVELOPMENTS

The artificial intelligence system described was run mostly under stringent program control. To make the system more flexible it is expedient to complement its intelligent and customized processors by a so-called instrumental processor. The function of this latter is to generate CS of variable architecture and structure, depending on the stage of the task being performed. The instrumental processor determines the number of atomic evaluators and their networking into a semantic net to optimize the search of a reference pattern for the image to be recognized and select the most efficient algorithm for the present stage. Thus, the intelligent processor sets the strategy, while the instrumental processor determines the tactics of recognition. Mathematical simulation of both processors utilized Petri nets.

REFERENCES

1. Y.Saito, S.Komatsu, H.Ohzu. Scale and rotation invariant real-time optical correlator using computer-generated hologram. - Optics Communication, 1983, v.47, No.1 pp.8-11.

2. D.Casasent, P.Psaltis. Positional, rotational, and scale invariant optical correlation. - Applied Optics. 1976, v.15 No.7, pp.1795-1800.

3. А.З.Дун, С.Ю. Маркин, Е.С.Невеженко и др. Исследование фотоэлектрического модулятора света в режиме обработки изображений. - Автометрия, 1982, с. 24 - 30, № 2.

4. Trends in speech recognition. Ed.W.Lea. Prentice-Hall,Inc.,Englewood Cliffs, N.J., 1980.

5. О.А.Бутаков, В.И.Островский, И.Л.Фадеев. Обработка изображений на ЭВМ. М., Радио и связь, 1987.

6. С.А.Майоров, Е.Ф.Очин, Ю.Ф.Романов. Оптические аналоговые вычислительные машины. Л., Энергоатомиздат, 1983.