COMPUTER TRANSCRIPTION OF MANUAL MORSE

Charles R. Blair
Silver Spring, Md.

A radio telegrapher can, by operating a key, turn a transmitter on or off for any desired period. Morse code is predicated on controlling these parameters of key position and the duration of the signal. A message is converted into a sequence of five elementary symbols: the dot, dash, intra-character space, inter-character space and inter-word space. In terms of an arbitrary unit of time, determined by the sender, a dot results from closing the key for one time-unit and a dash for three time-units. An intra-character space results from opening the key for one time-unit, an inter-character space for three time-units, and an inter-word space for seven time-units. All letters, numbers, punctuation marks and several brevity codes are represented by sequences of dots, dashes and intra-character spaces. (These code symbols are separated in a message by inter-character or inter-word spaces). Thus any message that can be transmitted by a sequence of characters can be broken down into a sequence of the five primary elements for transmission by radio-telegraph.

If manual Morse senders transmitted accurately, it would be easy to construct a device to measure the duration of each element, translate the elements into the appropriate sequence of code symbols, and transcribe this sequence into printed copy. Indeed, such devices have been constructed for transcribing signals sent by a machine. But, unfortunately, human operators are unable to control the duration of elements with sufficient precision. Experiments have shown, however, that durations representing the same element tend to form "clusters" about some typical value and, over a short span of text sent by a single operator, that these "clusters" are sufficiently distinct to permit assigning their component durations to specific elements.

Before a transmittal signal can be analyzed, it must be converted into digital form. To avoid developing a special purpose device, a routine was written in which the key is sampled by means of a conditional jump instruction. The time between successive samplings, regardless of the path taken by the routine, is programmed to be constant and therefore can be used as a real-time clock. A number of these samples are then averaged over the desired basic time unit to minimize the effects of contact "bounce". The sequence of durations determined in this fashion are stored in the order in which they are received.

From the sequence of stored durations, frequency distributions are formed for the key-open and key-closed elements. A separation line is arbitrally inserted between each pair of "clusters" on the basis of the a priori probabilities of the elements; then the line, in this region, which maximizes a "goodness of separation" statistic is determined iteratively. The set of dividing lines so calculated determine the range of durations which will be assigned to each element. In passing, it is interesting to note that this process is capable of classifying any set of scalars according to their tendency to cluster about some "typical" value. This association has been referred to as one-dimension pattern recognition.

Using the dividing lines calculated over a fairly short, and therefore presumably homogenous, stretch of text, each duration is assigned to a specific element. Sets of elements between inter-character spaces are then combined to form a number which, by table look-up, uniquely determines the character to be printed. Occasionally a combination of elements is obtained which does not correspond to any permissable character, indicating that a duration has been incorrectly assigned. Such errors usually result from mistaking an inter-character for an intra-character space, so the invalid character is split at the longest key open duration and retranscribed as two characters.

Experimental testing now in progress indicates that the process described is capable of transcribing manual Morse at least as well as an average human receiver.