

This document is also available in PDF<sup>1</sup> format.

**Purpose:** To introduce the concept of unsupervised training using VQ.

**Material:**

- Primarily LECTURE NOTES<sup>2</sup> on VQ, for reference also look at
- JA du Preez, “Spraakonafhanklike sprekerherkenning Chapter 5” (US MEng thesis)
- Makhoul, Roucos and Gish, “Vector Quantization ...”, Proceedings of the IEEE vol73 no 11, pp 1551 - 1588, Nov 1985
- Devijver & Kittler Chapter 11.

**General:** Quite often we are confronted with a situation where the data that we work with is made up of distinct, but unlabelled subclasses such as phonemes present in a word. It may even be that the subclasses are not all that distinct either, but we may still desire to divide the feature space into well-placed sub-regions for use in further modelling. If this is to be accomplished without explicit labelling of data (i.e. no supervisor), we are in the realm of clustering/VQ techniques. The term *VQ* is mostly used by the coding community, while *clustering* is more entrenched in PR circles. In the following, we will use the terms interchangeably. These algorithms attempt to subdivide the feature space into sub-regions sufficiently populated by feature vectors. These sub-regions may be investigated to determine if they approximate actual categories such as phonemes. Most often this is not the case, but they still remain useful.

**Topics:**

- This is a very wide and specialist field. Techniques abound, of which only a small subset is covered in the above material. The most frequently encountered is the so-called *K*-means algorithm (article pg. 1557). This is a special case of the dynamic clustering techniques of section 5.5.2 in the notes.
- As mentioned above, VQ and clustering are very closely related. Their original aims, however, were quite different and reflect the interests of the communities where they originated. With VQ an attempt is made to replace or quantise a continuous vector with a specific reference called a *codebook vector*, in such a way that minimal distortion is incurred according to the distance measure used. Clustering tries to discover the underlying structure or categories from which the feature space is made up. Two sides of the same coin, the first often emphasises the distances while the latter emphasises the models. Of course these are also closely related.
- It is interesting to note how the distance measure we use influences the shape of the clusters we discover! (A self-fulfilling prophecy, see section 5.2). Once again keep your eyes open for distance measures and their effects.
- An important topic concerns the matter of hard and soft decisions. *Hard* decisions (making an early decision and sticking to it) reduces computation considerably, but also gives worse results. Hierarchical clustering, such as the *K-means binary split* (pg. 1573 in article) is an example of this. An effective combination is often to start out with hard decisions for initialisation purposes, and then to refine it with a soft technique. A useful example is binary split followed by full *K*-means, see pg. 1579.
- Neural nets can also perform VQ; the Kohonen net is an example.
- These techniques are used extensively in speech coding. Other applications include discrete hidden Markov models for modelling speech, modelling speakers for recognition purposes, and also adapting the speech of a new speaker to that of the reference group to enhance speech recognition. We shall also see in later work that the supervised training of models for structured patterns, such as the HMM, also use conceptually very similar algorithms.
- A VQ approximation to some continuous feature space maps that space to a new discrete space. The mapping is from *N* dimensions to 1, and also from continuous to discrete. Because this mapping is not bound by classical pdf shapes, the VQ model is often considered to be non-parametric. Using the strict definition of “non-parametric”, this is, however, not true. The input features are represented / modelled by a series of parameters. In the simplest case these will be the codebook vectors, but it can also be a more complex representation. Some thought will show that modelling by means of codebook vectors is simply a special case of using multiple Gaussian PDFs. Using a diagonal Gaussian pdf with all variances equal to 0.5 and equal prior probabilities will result in exactly the same classification as that of a VQ codebook. From this viewpoint, it is a small step to generalise VQ to pdfs.

<sup>1</sup>[http://www.dsp.sun.ac.za/pr813/lectures/5\\_vq/5\\_vq.pdf](http://www.dsp.sun.ac.za/pr813/lectures/5_vq/5_vq.pdf)

<sup>2</sup>[http://www.dsp.sun.ac.za/pr813/lectures/lecture05/pr813\\_lecture05.pdf](http://www.dsp.sun.ac.za/pr813/lectures/lecture05/pr813_lecture05.pdf)

**Project:** (Working code to be ready by the next lecture, report will form part of the GMM task that will be handed out then.)

- Make sure that you understand the  $K$ -means algorithm. Read through all the given material, noting the different distance measures encountered. How do the measures affect the final results?
- Implement a speaker/face recogniser that represents each speaker/person with a VQ codebook. Classification is done by noting how far the unknown speech/image is from the codebook of each speaker/person. Compare with previous classification methods in terms of speed and accuracy.