



HAL
open science

EPG bidimensional data reduction

Noël Nguyen

► **To cite this version:**

Noël Nguyen. EPG bidimensional data reduction. International Journal of Language and Communication Disorders, 1995, 30, pp.175-182. 10.3109/13682829509082528 . hal-01392909

HAL Id: hal-01392909

<https://hal.science/hal-01392909v1>

Submitted on 15 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EPG Bidimensional Data Reduction

Noël Nguyen

*Laboratoire de Psycholinguistique, FPSE, Université de Genève
9 route de Drize, CH-1227 Genève, Suisse*

tel.: +41 22 705 97 41

fax: +41 22 300 14 82

e-mail: nnguyen@fapse.unige.ch

revised version submitted to the European Journal of Disorders of Communication

running title: EPG bidimensional data reduction

EPG Bidimensional Data Reduction

revised version submitted to the European Journal of Disorders of Communication

running title: EPG bidimensional data reduction

Abstract

A new method of electropalatographic (EPG) data reduction is proposed. It uses the so-called discrete cosine transform for representing an EPG pattern as a weighted sum of elementary “pictures”. We show that several of these pictures can be considered as being articulatorily relevant. The respective contributions of these pictures to the spatial arrangement of the linguo-palatal contacts are given by transform coefficients which are used here as a set of new EPG indices: the scaled sum of contacts index (SSU), the left-right asymmetry index (LRA), the alveolar-palatal index (APA), and the lateral-median index (LME). It is shown that the information extracted from an EPG pattern by means of these indices is maximized. Because of its robustness, our method could be usefully employed in studies on articulatory disorders.

key words: electropalatography, data reduction, articulatory parameter

1 Introduction

Electropalatography (EPG) is a technique for gathering data about the dynamic variations of the tongue-palate contacts in the production of speech (Hardcastle, Jones, Knight, Trudgeon & Calder, 1989; Marchal, 1988). In this domain, a great deal of attention has been recently devoted to the problem of data reduction (Fontdevila, Pallarès & Recasens, 1994; Hardcastle, Gibbon & Nicolaidis, 1991). EPG provides us with a vast amount of information which most often needs to be compressed. In the Reading EPG system for example, which we have used in the present work, tongue-palate contact patterns are recorded using a 62-electrode artificial plate, at a common sampling frequency of 200 Hz. When an experiment involves a large corpus and/or a large number of speakers, EPG recordings necessarily have to be transformed into a more manageable set of measurements. Moreover, besides its practical importance, EPG data reduction has also significant theoretical implications. Thus, previous studies on lingual movements in speech have led to the assumption that the tongue is characterized by a limited number of degrees of freedom. This hypothesis can be partially assessed by showing that it is possible to subdivide empirically the spatial arrangement of linguo-palatal contacts into a small number of independent components, by means of a statistical analysis (Nguyen, Marchal & Content, 1995). In this perspective, EPG data reduction methods can allow us not only to cut down the amount of information to be processed, but also to better understand how tongue movements are controlled.

Many methods of EPG data reduction have already been proposed (see Hardcastle Gibbon & Nicolaidis, 1991, for a review). To our knowledge however, no attempt has yet been made to apply to EPG data the techniques that are commonly used in the domain of image processing, although palatograms are images of a certain kind. In this paper, we present a new method for reducing EPG data based on such

a technique, namely the bidimensional discrete cosine transform.

2 Representing an EPG pattern as a weighted sum of basis pictures

In this work, it is assumed that an EPG pattern can be regarded as a binary image, in which activated electrodes are represented as black dots and unactivated electrodes as white dots. Our method consists in extracting the main features of this image by means of a two-dimensional spectral transform. It is well-known that 1-D spectral transforms, such as the discrete Fourier transform (DFT), are commonly applied to speech sounds, to determine how the acoustic energy is distributed on a frequency scale. In such transforms, the signal is converted into a weighted sum of orthogonal basis vectors (sampled sine and cosine functions in the case of the DFT). Analogously, an EPG image can be submitted to a 2-D transform which gives the possibility of characterizing the spatial arrangement of the tongue-palate contacts. In two-dimensional processing, the signal is also represented as a set of orthogonal components, which are called basis pictures (hereafter BPs). Interestingly, it is possible to interpret some of these pictures in articulatory terms, as will be shown below.

In our approach, EPG images are characterized using the so-called discrete cosine transform (DCT). This transform is easier to handle than the DFT, as it only involves real numbers. When applied to a two-dimensional array of 8×8 elements, such as an EPG image¹, the DCT is computed as follows (Clarke, 1985):

$$C_{pq} = C_0 \frac{\sqrt{2}}{8} \sum_{k=0}^7 \sum_{l=0}^7 f_{kl} \cos((2l+1)q \frac{\pi}{16}) \cos((2k+1)p \frac{\pi}{16}) \quad p = 0, \dots, 7; \quad q = 0, \dots, 7$$

$$C_0 = \begin{cases} \frac{1}{\sqrt{2}} & \text{if } p = q = 0 \\ 1 & \text{otherwise} \end{cases}$$

In this formula, f_{kl} is the value of the element located on row k , column l in the image field. The result is an array of spectral coefficients (C_{pq}), which represent the energy values respectively corresponding to 64 spectral components of increasing *spatial* frequencies. The low-frequency components reflect variations of intensity widely spread over the image surface. The high-frequency components are to be associated with more abrupt variations, which occur at more specific locations (e.g. edges or boundaries). The DC component C_{00} represents the average intensity level of the image elements.

As indicated above, the DCT may be viewed as a way of representing an image as a linear combination of BPs. In other words, this image can be reconstructed by simply summing the BPs, each weighted according to the individual spectral coefficients computed during the transform operation. Figure 1 depicts the first nine BPs, out of the 64 associated to an 8×8 image. The pictures are ordered as a function of spatial frequency.

— insert Figure 1 about here —

As shown in Figure 1, several of these pictures can be considered as being articulatorily relevant. For example, picture 1 is associated with the overall mean value of the image elements and, therefore, will bear a direct relation to the total number of linguopalatal contacts. Picture 2 is based on a subdivision into two parts in the horizontal dimension (columns 1–4 and columns 5–8). Thus, it may be used for modelling lateral asymmetries in the tongue-palate contact patterns. Picture 3 is

characterized by a zero-crossing in the vertical dimension (between the fourth and fifth rows), and may permit to locate the main concentration of contacts along the front-back axis. Picture 6 has two zero-crossings in the horizontal dimension, and will differentiate the EPG images depending on whether the tongue touches the palate in its medial part, or along the lateral sides.

Therefore, pictures 1, 2, 3 and 6 capture several characteristics in an EPG pattern that are relevant from an articulatory point of view. They can be considered as a set of EPG “features”, the presence or absence of each feature in a given pattern being indicated by the corresponding spectral coefficient.

One may ask what real weight is given to such features in the DCT. In other words, is it possible to show that these features predominate over other pictures clearly void of articulatory significance (e.g. picture 9)? To answer this question, a series of DCT were computed over a set of 2524 EPG patterns extracted from the EUR-ACCOR multichannel database (Marchal & Hardcastle, 1993). These data were gathered during the production of ten French sentences containing a large variety of vowels, semivowels, and consonants, by one female speaker (see Nguyen, Marchal & Content, 1995, for a complete description). For each pattern, 64 DCT coefficients were calculated. Overall root mean square values for the first 48 coefficients are displayed in Figure 2.

— insert Figure 2 about here —

Figure 2 gives rise to several remarks. First, the main part of the image energy resides in the lowest-order coefficients. Not surprisingly, this means that the configuration of the tongue-palate contacts most often formed a simple pattern, composed of one connected component only, with no gaps or irregular boundaries. Second, and more interestingly, Figure 2 shows that the three highest coefficients correspond

to BPs that have been identified above as having articulatory significance, namely picture 1, picture 3, and picture 6. Conversely, note that picture 2, that we have also interpreted in articulatory terms (lateral asymmetry), was associated with a rather small DCT coefficient. This might indicate that, for our speaker at least, the linguopalatal contact patterns were relatively symmetrical on average. Finally, one other coefficient, namely coefficient 8, appears to account for a rather significant part of the image energy. Figure 1 indicates that this coefficient might reflect the presence of a semi-circular occlusion, such as that involved in the production of an alveolar stop, for example (although this will be true only if this coefficient has a negative value, since a positive value will correspond to a semi-circular pattern oriented in the opposite direction). In general, it seems plausible to assume that high DCT coefficients are related to articulatorily-relevant dimensions in an EPG pattern.

3 Four new EPG indices

The results presented above suggest that four new EPG indices can be defined on the basis of the DCT general formula. The first index simply represents the scaled sum of the activated electrodes (SSU) in an EPG frame. The left-right asymmetry index (LRA) is a measure of the difference in the number of activated electrodes between the left and right sides of the palate. This index has a positive value when there are more contacts on the left, and a negative value in the opposite case. The alveolar-palatal index (APA) characterizes the arrangement of the tongue-palate contacts along the front-back dimension. Predominance of contacts in the alveolar region (first four rows) is expressed as a positive value, whereas a negative value indicates that the number of contacts is higher in the palatal region (last four rows). The lateral-median index (LME) indicates whether there are more linguopalatal

contacts along the lateral sides of the palate (positive values) or along the median line (negative values). These four indices correspond to the BPs 1, 2, 3 and 6 in Figure 1, respectively, and they are computed as follows:

$$\begin{aligned}
 \text{SSU} &= \frac{1}{8} \sum_{k=0}^7 \sum_{l=0}^7 f_{kl} \\
 \text{LRA} &= \frac{\sqrt{2}}{8} \sum_{k=0}^7 \sum_{l=0}^7 f_{kl} \cos\left((2l+1)\frac{\pi}{16}\right) \\
 \text{APA} &= \frac{\sqrt{2}}{8} \sum_{k=0}^7 \sum_{l=0}^7 f_{kl} \cos\left((2k+1)\frac{\pi}{16}\right) \\
 \text{LME} &= \frac{\sqrt{2}}{8} \sum_{k=0}^7 \sum_{l=0}^7 f_{kl} \cos\left((2l+1)3\frac{\pi}{16}\right)
 \end{aligned}$$

Figures 3 and 4 illustrate the kind of information our indices can provide, for some representative patterns extracted from the EUR-ACCOR database. Figure 3 shows linguo-palatal contact configurations at the onset of /k/ (upper panel) and at the end of closure for /l/ (lower panel), in the production of the nonword /akla/ by one female speaker. Besides each EPG pattern are displayed the four indices, each above the corresponding BP. The index values shown here are normalized, as they have been divided by the sum of the 64 DCT coefficient square values (i.e. the image total “energy”). Then, each BP has been weighted by the associated index value, to indicate how much this picture contributes to the EPG pattern.

— insert Figure 3 about here —

Figure 3 shows that the EPG pattern for /k/ was characterized by a moderate amount of contacts (13 activated electrodes, SSU = .45), a predominance of contacts along the lateral sides (LME = .2) and in the rearmost part of the palate (APA = −.56), and a slight right asymmetry (RLA = −.04). The EPG pattern for /l/ had one more activated electrode, as reflected by the slightly higher value of the SSU index (.47), it had more contacts on the right than on the left (LRA = −.17), and more contacts in the alveolar zone than in the palatal zone (APA = .22). Note that the APA index had a relatively small absolute value, given that lateral contacts

extended far toward the posterior edge of the palate. The measured ratio between the numbers of contacts along the lateral sides and along the median line was the same as for /k/ (LME = .2).

The tongue-palate contact patterns for /s/ in the production of /asa/ (upper panel), and for /ʃ/ in the production of /aʃa/ (lower panel), for the same speaker, are depicted in Figure 4. Each pattern was taken at the fricative point of maximum constriction.

— insert Figure 4 about here —

For the alveolar fricative, Figure 4 shows that the number of activated electrodes was relatively high (26, SSU = .64), and that the contacts were arranged rather symmetrically on both sides of the median line (LRA = $-.05$). The tongue was braced against the lateral edges of the palate (LME = .47), and the point of constriction was located in the front alveolar zone although there were linguo-palatal contacts along the entire length of the palate (APA = .25). The fricative /ʃ/ was characterized by a smaller number of contacts (20, SSU = .56), no left-right asymmetry (LRA = 0), and a mid-sagittal grooving of the tongue over the entire zone of linguopalatal contacts (LME = .44). The point of constriction was in the post-alveolar zone (APA = $-.02$).

4 Conclusion

The main advantage of the method proposed in this work resides in the fact that the DCT BPs are *completely orthogonal*. In other words, multiplying these BPs with one another always gives an entirely white image as a result². Therefore, the four indices that we have devised are related to dimensions which are completely independent from each other. In that sense, the information they provide us with

about the spatial configuration of linguo-palatal contacts is maximized.

Another interesting feature is that our indices are quite robust, and are defined in a way which does not depend on any speaker-specific characteristic. In this respect, our method could be usefully employed in studies on defective utterances, for which more classical EPG parameters might prove to be more difficult to measure.

Acknowledgements

This work was supported by the ESPRIT Working Group # 7098. Thanks are due to Paola Merlo for useful stylistic suggestions. We are also most grateful to Fiona Gibbon for helpful comments.

Bibliography

- Clarke, R.J. (1985). *Transform Coding of Images*. London: Academic Press.
- Fontdevila, J., Pallarès, M.D., & Recasens, D. (1994). The contact index method of electropalatographic data reduction. *Journal of Phonetics* **22**, 141–154.
- Hardcastle, W.J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* **19**, 251–266.
- Hardcastle, W.J., Jones, W., Knight, C., Trudgeon, A., & Calder, G. (1989). New developments in electropalatography: a state-of-the-art report. *Clinical Linguistics & Phonetics* **3**, 1–38.
- Marchal, A. (1988). *La Palatographie*. Paris: CNRS.
- Marchal, A., & Hardcastle, W.J. (1993). ACCOR: instrumentation and database for the cross-language study of coarticulation. *Language and Speech* **36**, 137–153.
- Nguyen, N., Marchal, A., & Content, A. (1995). Modeling tongue-palate contact patterns in the production of lingual consonants. *Submitted for publication*.

Footnotes

¹The Reading electropalate has 62 electrodes arranged in eight horizontal rows, with six electrodes only in the frontmost row. For the sake of simplicity however, an EPG “image” is considered here as having eight rows and eight columns (the two supernumary dots being always white).

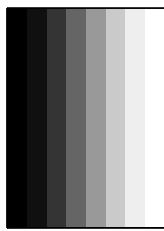
²In mathematical terms, this means that the product of the two corresponding matrices is a matrix only composed of zeros.

Figure captions

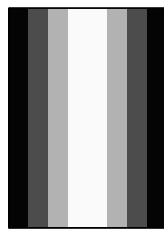
1. First nine basis pictures for the discrete cosine transform (after Clarke, 1985).
The pictures are ordered according to spatial frequency. White corresponds to numerical values comprised between -1 and 0 ; black corresponds to values between 0 and 1 .
2. Root mean square values of the DCT first 48 coefficients calculated over a set of 2524 EPG frames.
3. EPG patterns for /k/ onset (upper panel) and for /l/ release (lower panel) in the production of the sequence /akla/. The normalized values of the indices are indicated above the corresponding basis pictures. Each picture has been weighted by the associated index value.
4. EPG patterns for /s/ in /asa/ (upper panel) and for /f/ in /afa/, with the corresponding index normalized values.



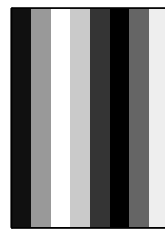
1



2



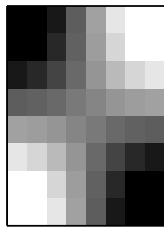
6



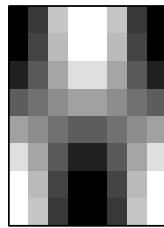
7



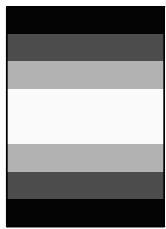
3



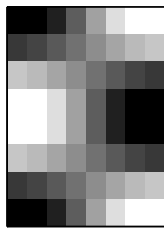
5



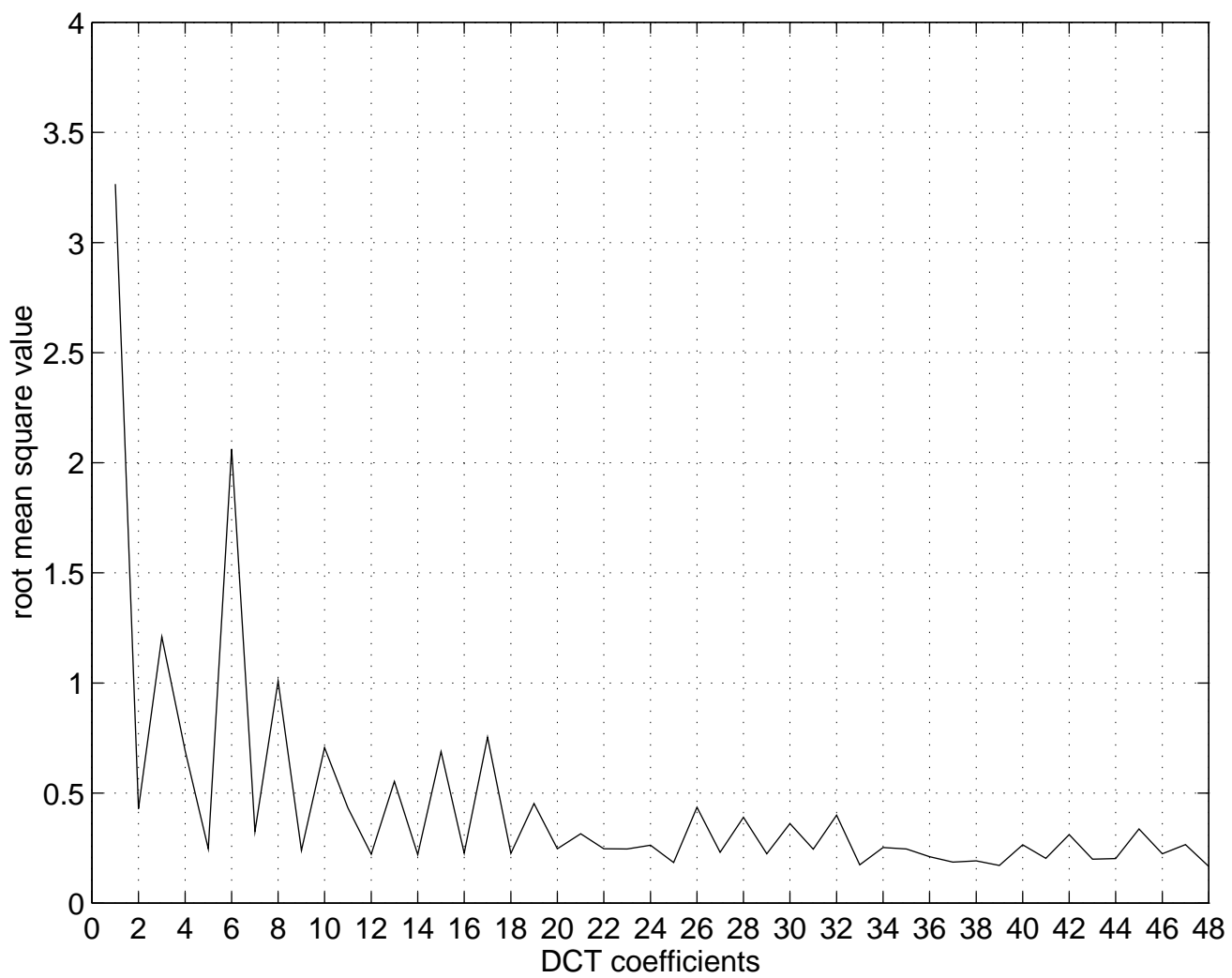
8

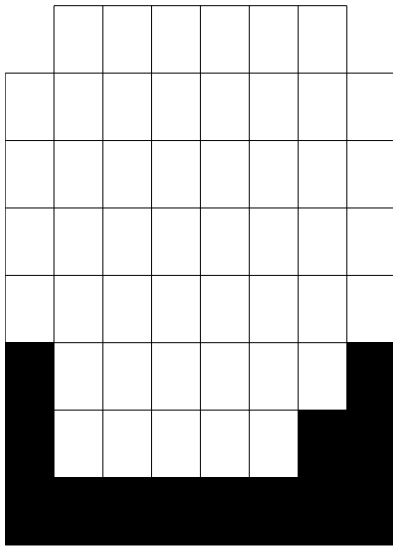


4

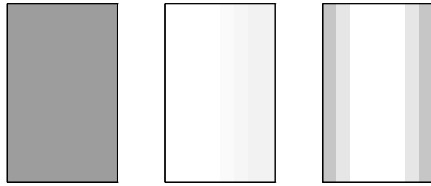


9

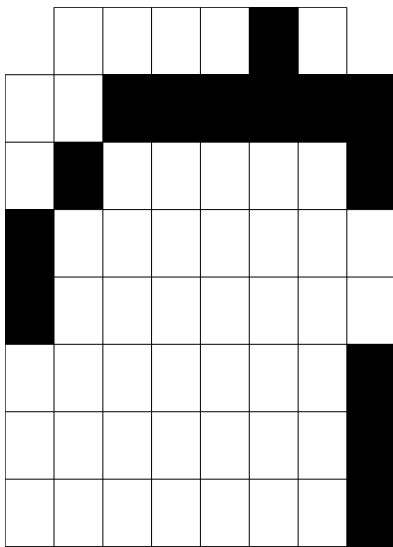
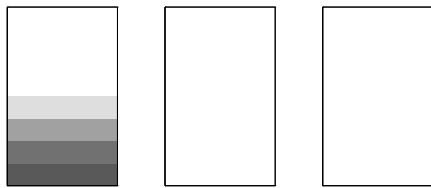




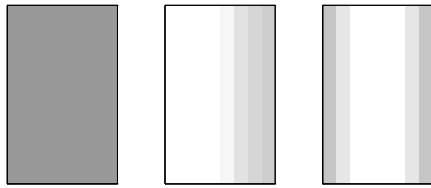
SSU = 0.45 LRA = -0.04 LME = 0.20



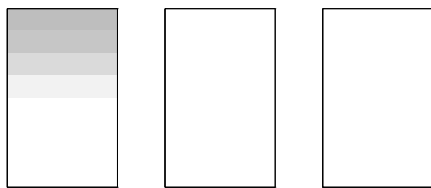
APA = -0.56

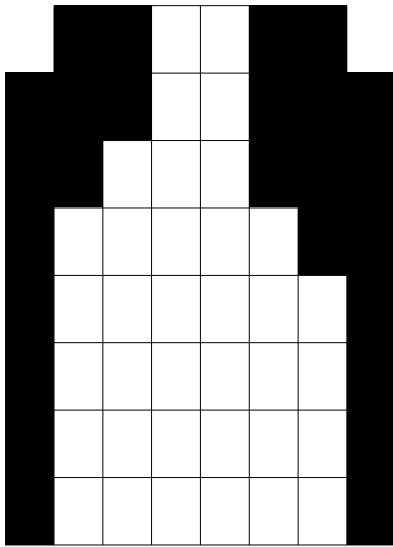


SSU = 0.47 LRA = -0.17 LME = 0.20

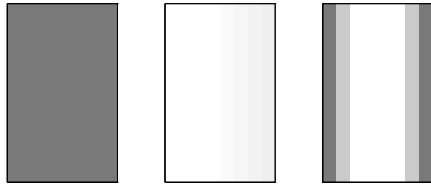


APA = 0.22

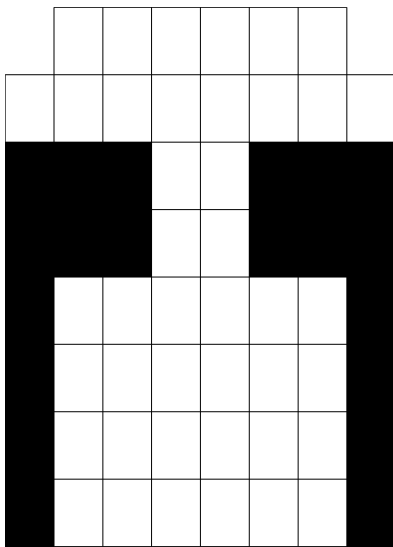
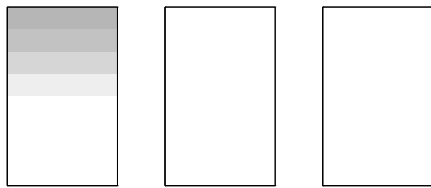




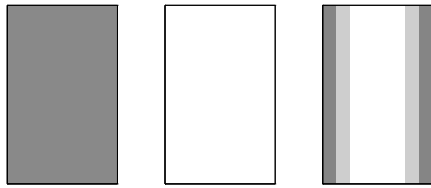
SSU = 0.64 LRA = -0.05 LME = 0.47



APA = 0.25



SSU = 0.56 LRA = 0.00 LME = 0.44



APA = -0.02

