Chapter 10

Experiments with real-world data: the acoustic-to-articulatory mapping problem

In this chapter, we apply our reconstruction algorithm to a version of the acoustic-to-articulatory mapping, a well-known mapping inversion problem of speech research. It is a complex, high-dimensional task that helps to further understand the performance of the algorithm. Before the description and discussion of the experimental results of section 10.2, we give some background of the problem and its significance for speech perception and automatic speech recognition (ASR) in section 10.1.

10.1 The acoustic-to-articulatory mapping problem

We describe the problem of articulatory inversion, its relation with the motor theory of speech perception, computational approaches for its solution and speech models incorporating articulatory information.

10.1.1 The problem

Broadly speaking, the acoustic-to-articulatory mapping problem consists of determining the vocal tract shape that produced a certain acoustic signal. The forward mapping, from a vocal tract shape or articulatory configuration to the acoustics, is univalued but nonlinear and many-to-one, which makes its inversion difficult (see fig. 10.1). The problem is also further complicated by the fact (among others) that the articulatory and

This chapter is partly based on references Carreira-Perpiñán and Renals (1999); Carreira-Perpiñán (2000b).

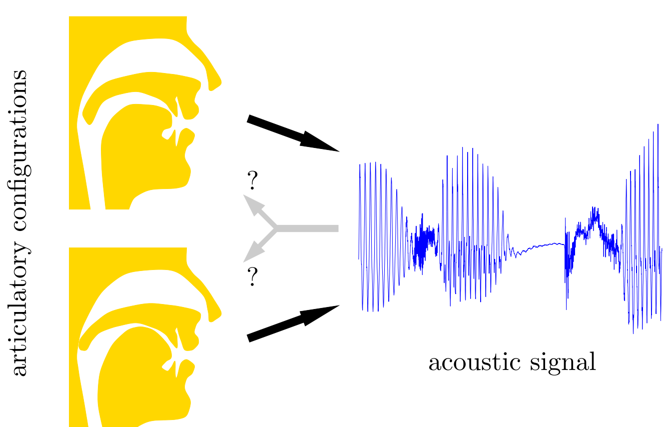

Figure 10.1: The acoustic-to-articulatory mapping problem: one vocal tract configuration produces a unique acoustic signal, but certain acoustic signals may be produced by multiple vocal tract configurations.

acoustic variables are not as clear cut as in, say, the robot arm problem of section 9.3. However, the problem is very important from both engineering and perceptual points of view.

10.1.1.1 Forward mapping: sound propagation in the vocal tract

The vocal tract acts as an acoustic filter that modifies the spectrum of the excitation signal at the glottis. From the point of view of articulatory synthesis one has to model the following elements (Fant, 1970; Flanagan, 1972; Schroeter and Sondhi, 1994):

Geometry of the vocal/nasal tract The vocal tract can be idealised as a straightened, nonuniform acoustic tube extending from the glottis ($x = 0$) to the lips ($x = L$) whose cross-section *area function* $A(x)$ varies continuously but slowly as a function of time. Information about its geometry may be obtained from X-ray measurements.

Wave propagation in the tract It can be described by *Webster's horn equation* (first derived by Bernoulli, Euler and Lagrange in the XVIII century). This is a second-order linear differential equation for the pressure (and volume velocity) as a function of x for a glottal signal and a given $A(x)$. Its solutions are plane waves, i.e., pressure and velocity are constant in a plane perpendicular to the tract axis. The equation is valid as long as the greatest cross-dimension of the tract is appreciably less than a wavelength, which means for frequencies smaller than 4 kHz. Nonlinear effects are important with turbulent flow (high Reynolds or Mach number), which happens through the vocal cords or through narrow constrictions as in fricatives[1]. The equation can be extended to account for effects of energy loss due to viscous friction, thermal conduction and yielding tract walls.

Sound sources and their interaction with the tract This requires a nonlinear model of the glottis (vocal cords).

Therefore the shape of the vocal tract is completely specified at any one time by the area function $A(x)$.

The **direct problem**, i.e., to determine the volume velocity and pressure of the air given $A(x)$ and certain other parameters, can be solved for any given boundary conditions at the lips and glottis. That is, the articulatory configuration plus the glottal source causally determine the acoustics. This allows speech synthesis from articulatory parameters, and is usually straightforward but computationally expensive. The **inverse problem**, i.e., to obtain articulatory information (in particular $A(x)$) from acoustic information extracted from the speech signal, does not have a unique solution: the transfer function of the vocal tract does not uniquely specify the area function[2] and so the acoustic signal at the lips does not either. There are two kinds of nonuniqueness:

- Different tract shapes may have or almost have the same transfer function, thus producing the same acoustic signal for the same given input at the glottis.
- The same acoustic signal may be produced by two different tract shapes with appropriate inputs at the glottis, i.e., changes in the source can compensate for certain changes of the tract transfer function.

Most models address only the former. The only way to deal with the nonuniqueness problem is to use constraints on the area function, particularly temporal continuity.

10.1.1.2 Articulatory variables and their properties

The articulatory configuration or vocal tract shape can be represented in various ways depending on the purpose of the inversion, but in any case it should give a reasonably complete description of the shape of the vocal tract. The area function $A(x)$ is such a complete description but for computational convenience a finite set of articulatory variables[3] is used. This can be achieved by discretising the area function or by using

[1] The main characteristics of speech sounds are (Ladefoged, 2000; Rabiner and Juang, 1993): *voicedness*, where the tract is excited by vibrating vocal cords (e.g. vowels, diphthongs); *frication*, where the tract is excited by turbulence due to flow through a narrow constriction (e.g. /s,z,ʃ,ʒ,f,v/); *plosiveness (stops)*, where the tract is excited by sudden release of pressure (e.g. /p,t,k,b,d,g/); *nasality*, where part or all of the flow is diverted into the nasal tract (e.g. /m,n,ŋ/); *silence*, where there is no excitation (e.g. during stop closures).

[2] The acoustic input impedance of the vocal tract at either end of the tract does uniquely specify the area function (for lossless vocal tracts only). But such information is not easily measurable.

[3] In the acoustic-to-articulatory mapping literature the representation of the vocal tract is usually called articulatory features or parameters or configuration or shape and the representation of the speech signal is usually called acoustic features. We will generally call them articulatory variables and acoustic variables in accordance with the rest of this thesis.

instead the positions of particular articulators or landmarks of articulators, such as the tongue tip, tongue body, jaw, velum, lip opening, etc. Ladefoged (1980) has suggested a set of 16 articulatory parameters which are necessary and sufficient to uniquely characterise all the sounds of every known language, but most studies use fewer than 10 and restrict them to the midsaggittal plane (the plane of fig. 10.1).

The articulators are masses accelerated with finite forces and occupy space, so they are subject to mechanical constraints. Various kinds of constraints have been proposed: static, which discard physically unrealisable articulatory configurations (e.g. 'tongue tip cannot go through the roof of the mouth'); dynamic, in that they change continuously and slowly with the time; and others such as those derived from the economy of skilled movements (Nelson, 1983), e.g. minimal muscle effort or work.

Data for the articulatory variables and their constraints can be derived from measurements with X-rays or EMA (section 7.10.5) or from articulatory models. An **articulatory model** is a geometric model of the vocal tract in terms of several parameters. For example, in the articulatory model of Mermelstein (1973) the parameters are the locations of tongue body centre, velum, tongue tip, jaw, lips and hyoid. The model allows the computation of the area function $A(x)$ that results from given values of the parameters; such area functions can then be used in Webster's horn equation. Articulatory models are aimed at representing mechanical constraints of the vocal tract and so they can be used to generate feasible vocal tract configurations (i.e., not all values of the parameters are allowed). The most favoured models are those of Mermelstein (1973), Coker (1976) and Maeda (1982). They are primarily based on (often two-dimensional) X-ray studies of the vocal tract. For the purpose of articulatory inversion one can use the articulatory parameters directly rather than the area functions.

Besides the squared reconstruction error, or root-mean-square (RMS) error, two other measures of quality of articulatory recovery are often used:

- Pearson product-moment correlation, which quantifies for a given articulator the similarity in shape between two trajectories regardless of magnitude, i.e., whether they rise and fall in synchrony:

$$r \stackrel{\text{def}}{=} \frac{\operatorname{cov}\{a,b\}}{\operatorname{stdev}\{a\}\operatorname{stdev}\{b\}} = \frac{\sum_{n=1}^{N}(a^{(n)}-\overline{a})(b^{(n)}-\overline{b})}{\sqrt{\left(\sum_{n=1}^{N}(a^{(n)}-\overline{a})^2\right)\left(\sum_{n=1}^{N}(b^{(n)}-\overline{b})^2\right)}} \in [-1,1] \tag{10.1}$$

  for two sequences $\{a^{(n)}\}_{n=1}^{N}$ and $\{b^{(n)}\}_{n=1}^{N}$ of means $\overline{a}$ and $\overline{b}$, respectively.

- In the context of automatic speech recognition (ASR), some measure of articulatory gesture or phoneme recognition, such as a phone classification score.

10.1.1.3 Acoustic variables and their properties

The raw acoustic signal as a function of time is not convenient because of its high sampling rate (normally around 20 kHz) and variability (due to inter- and intraspeaker variability, noise and coarticulation). Depending on the problem, other representations (Rabiner and Juang, 1993; Gold and Morgan, 2000) are used that result in a vector time series with a rate of the order of 100 Hz (closer to that of the articulators). In decreasing order of closeness to the vocal tract shape:

Formants The formants are the resonances of the vocal tract and are therefore very closely related to its shape, changing slowly with time and showing relatively simple phonemic transitions. Besides, they are quite robust to noise. However, they cannot be generally used: they are not always visible in the spectrum (e.g. when a narrow constriction decouples the rear cavity, as in fricatives) and they are difficult to extract reliably. The formants have been often used in studies of the acoustic-to-articulatory mapping for vowels.

Linear predictive coding (LPC) performs spectral analysis on speech frames with an all-pole filter. It provides a good approximation to the vocal tract spectral envelope for voiced speech and achieves a reasonable source-vocal tract separation. It is less effective for unvoiced and transient regions of speech. Other variations of LPC are *line spectral frequencies* (LSF) and *line spectral pairs* (LSP).

Filter banks The speech signal is passed through a bank of several independent but overlapping bandpass filters collectively spanning the frequency range of interest. Thus, the output of each filter is a short-time spectral representation of the signal at the filter's centre frequency at the current time frame.

Auditory-based cepstral representations A smoothed short-term spectrum is derived from a filter bank that has been designed according to some model of the auditory system. The features are decorrelated with a linear transformation which also separates out pitch, spectral detail and spectral tilt. The most common variants are the *mel cepstrum*[4] (MFCC) and *perceptual linear prediction* (PLP) (Hermansky, 1990), which provide very similar features; a more recent proposal are *modulation-filtered spectrogram* (MSG) features (Kingsbury et al., 1998), developed for speech recognition robust to acoustic interference such as additive noise and reverberation. In addition to the cepstral coefficients, estimates of their velocities and accelerations are often used to account for dynamic features of speech. Most current ASR systems use this representation. However, cepstra are sensitive to noise (because of the logarithmic compression and subsequent spread over all features by the linear transformation), to coarticulation and speaker-dependent; and they present complex transitions and discontinuities where the excitation changes.

Thus, while the raw acoustic waveform is continuous in time, the acoustic variables generally are not, even for representations like LSPs which are closer to the formants. And articulatory trajectories that differ only slightly can result in very different acoustic utterances. Deng et al. (1997) mention a good, well-known example: stop epenthesis after nasals. This occurs when an extra silence and burst are introduced in the acoustic signal due to variation in timing between adjacent velic and oral closures. For example, the realisation of the word "princess" as [printses] or [prinses] depends on whether the velum is raised before or after the release of the alveolar stop, and the amount of desynchronisation varies continuously (as observed in articulatory data). Accounting for this in the acoustic domain is far more difficult than in the articulatory domain, requiring either to assume an extra stop phoneme for the word in question or to extend the acoustic model of the nasal to include the epenthetic stop.

In many of the computational approaches for the acoustic-to-articulatory mapping problem, an **acoustic distance** is required, i.e., a distance between acoustic vectors. Many such distances have been proposed in the speech literature, often quite complex to make them more relevant to human perception (distortion measures). For filterbank and cepstral vectors, an L_1, L_2 or covariance weighted spectral difference is often used; for LPC coefficients, the likelihood ratio distance is more appropriate (Rabiner and Juang, 1993, chapter 4). Likewise, an **articulatory distance** is required, and several have been proposed, with the Euclidean one being often used.

10.1.1.4 Example of the nonuniqueness

Many examples of the nonuniqueness of the acoustic-to-articulatory mapping have been given in the literature, resulting from both articulatory models and human experiments (Schroeter and Sondhi, 1994). A familiar demonstration are ventriloquists, who can produce intelligible speech without moving the lips. Another often-cited example is that of the approximant consonant /ɹ/ of American English (the 'r' as in 'beaker', 'perk', 'rod' or 'street') (Westbury et al., 1998; Espy-Wilson et al., 2000). Speakers of rhotic dialects of American English use many different articulatory configurations for /ɹ/, which are all acoustically characterised by an extremely low frequency of the third formant (often close to that of the second formant). These configurations expose an ante/sub-lingual cavity and involve three constrictions: in the pharynx, along the palate and at the lips. The configurations differ most in the palatal constriction and have traditionally been divided into contrasting categories of *retroflex* (tongue tip raised, tongue dorsum lowered) and *bunched* (tongue dorsum raised, tongue tip lowered), but there really seems to exist a continuum between them. These different configurations occur both within and across speakers: some speakers may use one type of configuration exclusively while others switch between two or three different types in different phonetic contexts and according to prosodic variables.

Computational models have also confirmed the nonuniqueness of the acoustic-to-articulatory mapping. Atal et al. (1978) found articulatory regions (fibers) that map onto a single acoustic point by linearising the forward mapping in a small neighbourhood and extending it in small steps. They found that many sounds can be produced by many different vocal tract shapes.

Finally, from a theoretical standpoint, the nonuniqueness appears for lossless vocal tracts with fixed boundary conditions: the area functions $A(x)$ and $1/A(L-x)$ (where L is the vocal tract length) produce the same transfer function. For lossy vocal tracts, it is not clear whether the nonuniqueness remains, but practically more than one area function produce very similar transfer functions.

[4]The (complex) cepstrum of a signal is the Fourier transform of the log of the signal spectrum (which is itself the Fourier transform of the signal).

10.1.1.5 Coarticulation

Coarticulation broadly refers to the fact that a phonological segment is not realised identically in all contexts, but often apparently varies to become more like an adjacent or nearby segment (Hardcastle and Hewlett, 1999). It can be anticipatory (a later segment influences an earlier one) or carryover (vice versa). For example, the English phoneme /k/ is articulated further forward on the palate before a front vowel ([k̟iː] ‘key’) and further back before a back vowel ([k̠ɔː] ‘caw’); and will have a lip position influenced by the following vowel (e.g. rounded in [k̠ʷɔː] ‘caw’). As another example, in velopharyngeal coarticulation nasality spreads from a consonant to a neighbouring vowel: compare the /a/ in /as/ and /an/.

The reason for coarticulation is that the vocal tract cannot move from one target configuration to the next one instantaneously, so instead of keeping each phoneme as an invariant articulation and then slowly moving to the next, the articulators follow a faster, more graceful trajectory. The higher the coarticulation the more fluent the sequence and the more difficult it is to isolate individual phonemes. The same happens in handwriting.

Coarticulation is thought to have advantages for perception: spreading the effect of a phoneme to a larger interval makes it more likely to be spotted and several phonemes may be processed in parallel. Thus, coarticulation could be an adaptation of the human communication system to maximise the transmission rate at its bottleneck: the slow-moving articulators. However, coarticulation makes each acoustic unit depend heavily on its context, which makes speech recognition difficult. Dealing with coarticulation should be much easier in its native, articulatory domain than in the acoustic one.

There is a parallelism in terms of planning between an utterance and a robot arm trajectory in that targets must be met: in the utterance, these are acoustic targets given by the phonemic transcription of the utterance, while in the robot arm these are physical locations through which the end-effector must pass (perhaps avoiding obstacles). The targets are given in the acoustic or work space and result in corresponding targets in the articulatory space. However, in speech the articulatory targets do not necessarily have to be fully realised for speech to be intelligible, particularly in fast speech styles, as shown by coarticulation.

10.1.1.6 Critical vs non-critical articulators (“don’t care” values)

The concept of critical articulators refers to the fact that, for a given production, the movement of a small subset of articulators is crucial, while the movement of the rest is not. For example, the acoustics, particularly the formants, are more sensitive to place and degree of constriction than to the rest of the area function. One reason for this is that the coefficients of Webster’s horn equation are functions of the logarithm of the area function. Recasens’ work on coarticulation in Catalan (reviewed in Hardcastle and Hewlett, 1999, chapters 2 and 4) showed that the more an articulator is involved in producing a consonant, the less susceptible it is to coarticulatory influences from adjacent vowels.

Papcun et al. (1992) demonstrated empirically that critical articulators are less susceptible to coarticulation and have a greater range of variation than noncritical articulators, which are freer to vary or play along. They used real articulatory data (recorded with an X-ray microbeam) and trained a neural net to learn the mapping from acoustics[5] (bark scaled FFT bins) to articulators’ positions (lower lip, tongue tip and tongue dorsum) for the English stop consonants /p,b,t,d,k,g/. The critical articulators were the lower lip for bilabials (/p,b/), the tongue tip for alveolars (/t,d/) and the tongue dorsum for velars (/k,g/). By comparing the articulatory trajectories inferred by the neural net with the original ones, they found good correlation ($r \approx 0.9$) for the critical articulators of each consonant type and bad correlation ($r = 0.19$ to 0.78) for the noncritical ones, but higher RMS error for the critical than for the noncritical ones. They hypothesise that intra-speaker variability of non-critical articulators could be caused by principled differences (such as differences in phrasal stress) or considered as noise; while inter-speaker variability could result from the fact that each speaker has acquired idiosyncratic patterns of noncritical articulator movement but shares patterns of critical articulator movement with other speakers.

Another case of “don’t care” values occurs during silence intervals in the speech (e.g. during stop closures), in which the output spectrum is similar to that of the background noise and does not contain any information regarding the shape of the vocal tract.

“Don’t care” values also occur in the robot arm problem (section 9.3). For example, Jordan (1990) considers the case of a robot with two arms: if at some moment there is only one target to manipulate, the arm which is not manipulating the target is free to move; but it may move towards a future target to anticipate a movement and so make the transition to the next configuration easier (faster, requiring less energy, etc.). Coarticulation

[5] Rather than using as input the acoustic vector of a single frame, they used a sequence of 25 frames (around 200 ms), which they called *context frame*.

is the analog case in speech production. The same phenomena should occur if the vocal tract is extended to include facial features (section 7.10.4).

10.1.1.7 Significance for speech processing

The constraints of the speech production system (such as slow, continuous dynamics, coarticulation, identification of critical articulators and speaker-dependent characteristics such as vocal tract length) can be better represented by an articulatory representation of speech than by an acoustic one. An articulatory representation should be useful for:

- Speech recognition: traditional speech recognisers have problems with nasals, voiced and unvoiced stops and voiced and unvoiced fricatives, for which spectrograms are very similar, discriminatory features concentrate on very few transitionary frames and the spectrum between transitionary frames is ambiguous or useless. They also have problems with the effects of prosody and coarticulation, which they partially alleviate by using context-sensitive models (e.g. diphones), but at the cost of using many parameters.

  The addition of articulatory features to acoustic features in HMMs has been shown to increase recognition performance over acoustic features alone (Zlokarnik, 1995a). We review some more sophisticated models using production information in section 10.1.4.

- Coding and text-to-speech synthesis: an articulatory representation has lower requirements than an acoustic one for transmission rate and storage because of the slow dynamics of the articulators. Articulatory trajectories are also closer to a phonetic transcription.

- Visualisation of vocal tract features and training aids for the deaf, etc. (as in chapter 5 with the EPG data).

In summary, some aspects of speech processing should be simpler in the articulatory domain. Since usually only the acoustic signal is readily available, articulatory inversion becomes necessary.

10.1.2 The motor theory of speech perception

Irrespectively of any mathematical or engineering approach, how the human brain may perform articulatory inversion is unknown[6]. We briefly review the well-known motor theory of speech and note an interpretation in terms of latent variables.

The basic motivation for speech theory is that people can both perceive speech and produce speech. It seems unparsimonious to assume that the speaker-listener employs two entirely separate processes, one for encoding language and the other for decoding it. A simpler assumption is that there is only one process with appropriate links between sensory and motor components. Speech is then assumed to be perceived by processes that are involved not only in auditory perception but also in speech production.

The motor theory (Liberman et al., 1967; Liberman and Mattingly, 1985) was originally proposed to try to account for perceptual invariance in the face of highly variable, context-dependent, acoustic cues. Experiments show that there is typically lack of correspondence between acoustic cue and perceived phoneme, which rules out the use of acoustic cues as perceptual primitives. For example, coarticulation effects, the fact that any particular acoustic segment will likely to be cueing more than one phoneme at a time. In several cases it appears that perception mirrors articulation more closely than sound. This supports the assumption that the listener uses the inconstant sound as a basis for finding his way back to the articulatory gestures that produced it.

The motor theory makes two basic claims: (1) the existence of an invariant motor code of phonetic gestures shared by speech perception and production; and (2) the existence of an innate, specialised module in the brain responsible for the translation between phonetic gestures and acoustic patterns. Let us examine these claims in more detail.

The existence of a motor code The objects of speech perception are the intended phonetic gestures of the speaker, represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations. These gestural commands are the physical reality underlying the traditional phonetic notions (such as tongue backing, lip rounding or jaw raising) that provide the

[6] Likewise, how the brain solves other motor problems, like arm control, is the subject of active research (for a recent review see Wolpert and Ghahramani, 2000). Concepts similar to those proposed by the motor theory appear there too, such as postulated motor primitives or forward and inverse internal models.

basis for phonetic categories. They are the elementary events of speech production and perception. Phonetic segments are simply groups of one or more of these elementary events; thus [b] consists of a labial stop gesture and [m] of that same gesture combined with a velum-lowering gesture. Phonologically, of course, the gestures themselves must be viewed as groups of features, such as labial, stop or nasal, but these features are attributes of the gestural events, not events as such. To perceive an utterance, then, is to perceive a specific pattern of intended gestures (more or less altered due to coarticulation and other effects).

As a detailed example of how a phoneme would be broken down, from a production point of view, into a sequence of possibly overlapping subphonemic elements, consider the articulation of /b/ (Liberman et al., 1967):

1. Closing and opening the upper vocal tract in such a way as to produce the manner feature characteristic of the stop consonants.
2. Closing and opening the vocal tract specifically at the lips, thus producing the placed feature termed bilabiality.
3. Closing the velum to provide the feature of orality.
4. Starting vocal fold vibration simultaneously with the opening of the lips, appropriately to the feature of voicing.

Other phonemes could be described using these and other gestures:

/p/ has features 1, 2 and 3 but differs in 4 in that vocal fold vibration begins some 50 or 60 milliseconds after opening the lips.

/m/ has features 1, 2 and 4 but differs in 3 in that the velum hangs open to produce the feature of nasality.

/d/ has features 1, 3 and 4 but differs in place of articulation.

Therefore, a phonetic, or motor, gesture can be defined as a class of movements by one or more articulators that results in a particular, linguistically significant deformation over time of the vocal tract. A gesture may be effected by several articulators for several reasons:

- A gesture may require the collaboration of several articulators. For example, lip rounding is a collaboration of the lower lip, the upper lip and the jaw.
- A single articulator may participate in the execution of two different gestures at the same time. For example, the lips may be simultaneously rounding and closing in the production of a labial stop followed by a rounded vowel, as in [bu].
- Prosody effects. For example, producing a stressed syllable requires a greater displacement of some or all of the active articulators than when producing an unstressed one.
- Linguistically irrelevant factors, notably speaking rate, affect the trajectory and phasing of the component movements.

The existence of a specialised module for the interface between speech perception and speech production The existence of a motor code implies the existence of an intimate link between speech perception and speech production. In the motor theory, this link is innate, not learned, and is implemented by a specialised module of the brain. Thus, perception of the gestures occurs in a specialised mode different from the auditory mode.

Computation of the phonetic gestures from the acoustic signal by a cognitive process does not seem reasonable. This justifies the need in the motor theory for a modular account of linguistic perception and the assumption of the existence of a special-purpose computational device that relates gestural properties to acoustic patterns. The conversion from acoustic signal to gesture (i.e., a form of articulatory inversion) is done automatically, so that listeners perceive phonetic structures without mediation by, or translation from, the auditory appearances that the sounds might, on purely psychoacoustic grounds, be expected to have.

The motor theory assumes that adaptations of the motor system for controlling the organs of the vocal tract took precedence in the evolution of speech over the development of a perceiving system. These adaptations made possible not only to produce phonetic gestures, but also to coarticulate them so that they could be produced rapidly. A perceiving system, specialised to take account of the complex acoustic consequences, developed concomitantly.

As biological basis for this specialised module, the motor theory proposes the existence of several neural networks—those that supply control signals to the articulators and those that process incoming neural patterns from the ear—with overlapping activity, so that information is correlated by these networks and passed through them in either direction.

10.1.2.1 Problems of the motor theory

The motor theory looks for a hidden representation of the speech message in terms of articulatory gestures, with information flowing in both directions (articulators $\rightarrow$ acoustics and acoustics $\rightarrow$ articulators), passing through the gestural representation. However, the idea of a gestural representation runs into a number of problems. A major shortcoming of the theory is the difficulty of rigorously defining, in physical terms, a particular gesture, due to the complications posed by coarticulation and other factors. This makes the motor gestures hardly more satisfactory as perceptual primitives than the acoustic cues. Further, categorising one group of the infinite number of possible articulatory movements as lip rounding and another as lip closure is entirely a priori. Besides, experiments in language acquisition in newborns have showed that structures of speech perception occur well before those of production.

Thus, while the mere ability of humans to listen and speak suggests that some sort of representation of the speech message must exist in the brain, it does not seem plausible that this representation is in terms of articulatory gestures, as the motor theory assumes. Several recent papers debating whether speech is controlled by auditory-acoustic goals or by articulatory goals appear in the *Journal of the Acoustic Society of America* 99(3):1680–1741 (March 1996); a summary is given by McGowan and Faber (1996).

There exist other feature-based theories, e.g. in phonology. In the theory of articulatory phonology of Browman and Goldstein (1992), the basic units of phonological contrast are called gestures and are also abstract characterisations of articulatory events, each with an intrinsic duration. Utterances are modelled as organised patterns of gestures, called constellations, in which gestural units may overlap in time. Thus, utterances differ from one another in the particular set of gestures they use or in how those gestures are organised, and the same gesture may have different acoustic consequences, depending on other concurrent gestures. The patterns of overlapping organisation can be used to account for several types of phonological variation, including coarticulation. Again, a listener must have a mechanism to recover the underlying gestures from the varying acoustics.

10.1.2.2 A latent-variable view

Regardless of its biological validity, the motor theory and other feature-based theories like that of Browman and Goldstein (1992) are computationally attractive and have been used as the motivation for several ASR approaches, some of which we have described in this thesis (e.g. Papcun et al., 1992; Erler and Freeman, 1996; Richards and Bridle, 1999). We point out here that the motor theory can be formulated as a latent variable model. Let us imagine the existence of a more abstract representation, neither expressed in terms of acoustic cues nor of articulatory gestures, and whose biological basis would reside in neural networks with bidirectional links with both the auditory and the articulatory systems. This could be implemented with a latent variable model trained in an unsupervised way with both acoustic and articulatory data (the latent variables would not necessarily be interpretable as a neural code); and the links between all three domains (acoustic, articulatory and the hidden representation) would be implemented by conditional probability rules.

10.1.3 Computational approaches

Webster's horn equation allows the computation of the acoustic signal resulting from a given area function for some glottal excitation, i.e., the articulatory-to-acoustic (forward) mapping, but not its inverse. Decades of research have been dedicated to computing this inversion. Levinson and Schmidt (1983) started their paper saying that

> The direct determination from a speech signal of the corresponding articulatory parameters, such as area functions or other representations of vocal tract shape, is a long standing problem in speech research.

Almost 20 years later, the problem of articulatory inversion is still unresolved, particularly for unvoiced sounds. Much of the work on it has been reviewed by Schroeter and Sondhi (1994), so we restrict ourselves here to the most important approaches as well as some of the more recent ones. Further review material can be found in some of the references cited in this section.

Many articulatory inversion methods can be used with the *analysis-by-synthesis* procedure, which is an optimisation closed loop in which the spectrum of the synthesised speech is compared to the real one at consecutive speech frames. For each frame, an optimisation procedure tries to minimise an acoustic distance between the two signals. In other words: take a starting articulator configuration $\mathbf{x}_0$; compute the forward mapping $\mathbf{g}(\mathbf{x}_0)$; backpropagate the error (acoustic distance) between the original speech $\mathbf{y}$ and the synthesised speech $\mathbf{g}(\mathbf{x}_0)$ to obtain an improved articulatory vector $\mathbf{x}_1$; iterate till convergence. The optimisation is initialised with an articulatory vector $\mathbf{x}_0$ obtained by some articulatory inversion method (e.g. from a codebook), which should be good to avoid local minima of the distance—which, in fact, is the major problem of this framework. Also, if the starting articulatory vector was good enough, one could avoid the optimisation loop altogether. Analysis-by-synthesis methods usually include as main parts an articulatory model (which describes the geometry of the oral cavity), an articulatory synthesiser (vocal tract model that simulates the physics of sound generation in that cavity), an optimisation algorithm with an error measure, and a spectral estimation algorithm (acoustic variables).

Three approaches to the acoustic-to-articulatory mapping are particularly important:

Dynamic programming search in a large articulatory codebook The use of articulatory codebooks was introduced by Larar, Schroeter, and Sondhi (1988) and its search by dynamic programming by Schroeter and Sondhi (1989). Earlier work by Atal et al. (1978) contains precursory ideas for codebooks. An articulatory codebook is a fixed, very large ($M > 100\,000$ entries) table of M vocal tract shapes (obtained either from measurements with X-ray, EMA, etc. or by sampling an articulatory model) and their respective acoustic output. It disregards glottal excitation. The whole codebook is scanned at every speech frame and the optimal path found by dynamic programming[7] to minimise a cost function of the form (a particular case of eq. (7.2)):

$$\underbrace{\lambda \sum_{n=1}^{N-1} \left\| \hat{\mathbf{x}}^{(n+1)} - \hat{\mathbf{x}}^{(n)} \right\|^2}_{\mathscr{C}} + \underbrace{\sum_{n=1}^{N} \left\| \hat{\mathbf{y}}^{(n)} - \mathbf{y}^{(n)} \right\|^2}_{\mathscr{F}} \tag{10.2}$$

where $\mathbf{y}$ represent original acoustic vectors, $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ represent codebook vectors (with $\hat{\mathbf{y}} \stackrel{\text{def}}{=} \mathbf{g}(\hat{\mathbf{x}})$), $\mathbf{g}$ is the known[8] articulatory-to-acoustic mapping, the $\mathscr{C}$ term is the continuity constraint (applied solely to the articulatory variables) and the $\mathscr{F}$ term is the forward mapping constraint (applied to all variables, but the articulatory ones cancel out). Thus, the $\mathscr{C}$ term is based on an articulatory distance (typically Euclidean) while the $\mathscr{F}$ term is based on an acoustic distance (of which many variations exist in the speech literature; section 10.1.1.3). Another constraint that has often been used for the articulators (e.g. Shirai and Kobayashi, 1986; Sorokin, 1992; Yehia and Itakura, 1996) is muscle effort, given by a quadratic function of the displacement of the articulators (see section 7.5):

$$\sum_{d=1}^{D_1} c_d (x_d^{(n)} - \overline{x}_d)^2 \tag{10.3}$$

where $\overline{x}_d$ is the equilibrium position of the dth articulator and c_d is a coefficient of tissue elasticity. It can be extended to a Mahalanobis distance or some other distance between the articulatory vector $\mathbf{x}^{(n)}$ and the constant equilibrium vector $\overline{\mathbf{x}}$. Determination of the acoustic-articulatory cost weight λ has been tried in various heuristic ways (Schroeter and Sondhi, 1994).

Although there are heuristic techniques to speed up the search (such as selecting at each frame only the M' best acoustic fits, with M' of the order of $1\,000$), since M is so large, the dynamic programming search is very slow and must be limited to around $N = 20$ frames (= 200 ms of speech for a frame shift of 10 ms), having a computational complexity of $\mathcal{O}(NM^2)$.

[7] This use of dynamic programming must be distinguished from the *dynamic time warping* algorithm (Rabiner and Juang, 1993, pp. 200–240), in which dynamic programming and various forms of constraints are applied in pattern comparison for speech recognition purely in the acoustic domain. There, the goal is to align in time and normalise the distance between *two* speech patterns (sequences of spectral vectors) of possibly different durations (number of vectors in the sequence). The constraints try to ensure that there is a proper time alignment by avoiding time reversal (monotonicity constraint) and obtaining a smooth alignment path between the starting point and the end point, both of which are fixed and given by the sequences' length (continuity, slope, endpoint and global path constraints).

[8] "Known" means that either it can be derived analytically from a physical model or it can be reliably approximated from data for the inputs and outputs.

Dynamic programming search of codebooks is currently the most accurate method of articulatory inversion. Its disadvantages are the large size of the codebook, which consequently takes a large storage and results in a slow search (of the order of 300× slower than real time for a 100 000-entry codebook in a 40-MFLOP computer; Schroeter and Sondhi, 1994); and the difficulty of constructing a good codebook. Generating the codebook is a careful process that requires:

- A method to obtain training vectors that adequately span both the articulatory and the acoustic spaces. This can be done in two ways:
  - From measurements (e.g. X-rays): the difficulty is to thoroughly span the articulatory space, since such measurements are a finite collection of one-dimensional trajectories designed in the acoustic space.
  - By finely sampling an articulatory model: this needs to eliminate unfeasible shapes. Another way is to interpolate along an articulatory trajectory determined by some predefined, "root" shapes, but this usually leaves areas of the articulatory space uncovered.
- A method to cluster the training vectors. Quantisation is attained by a clustering algorithm, which can be complicated and time-consuming depending on the acoustic distance used. For example, an intermediate point between two sample points may be associated with an unfeasible articulatory configuration and so standard k-means does not work.
- A definition of distances in both spaces, articulatory and acoustic.

Global parametric mappings with more or less sophisticated architectures, such as polynomials, radial basis functions or neural networks (trained with codebook vectors or measured data). Neural networks are faster and more compact than codebooks, but produce worse mappings, which is not surprising in view of the nonuniqueness of the mapping (section 7.3.4). An example using neural networks (though not the first one) was that of Papcun et al. (1992), described in section 10.1.1.6.

Methods based on local acoustic-to-articulatory mappings These methods try to split the acoustic space into regions such that each region maps one-to-one to a corresponding region in the articulatory space (branch determination step). Inside each region, a function approximator is used, trained in a supervised way using the data (or codebook) vectors that fall in the region. In particular, Rahim et al. (1993) (building on unpublished work by Parthasarathy and Sondhi described by Schroeter and Sondhi, 1994) cluster the training data as described in section 7.11.2.1 resulting in $N_{\mathbf{y}} = 32$ acoustic clusters each containing $N_{\mathbf{x}} = 4$ articulatory subclusters. They then fit a different MLP with 26 hidden units in each of the $N_{\mathbf{x}}N_{\mathbf{y}} = 128$ clusters. At the dynamic programming search stage (which is carried out every 15 pitch periods), first the centroids of the $N_{\mathbf{x}} = 4$ clusters are searched for the one closest to a given acoustic vector; then the $N_{\mathbf{y}} = 32$ mappings for the selected acoustic cluster are used to compute $N_{\mathbf{y}}$ mapped articulatory vectors, which are declared as possible candidates. Rahim et al. claim that the ensemble is 20 times more efficient than a codebook method both in memory and lookup time with results of similar quality. An extra advantage over the codebook is that, after having been trained using the codebook, the MLPs can be bootstrapped from natural speech. The disadvantages of this method were mentioned in section 7.11.2.1:

- There is no guarantee that the local mappings are one-to-one inside every region, and determining the regions is difficult in high dimensions.
- $N_{\mathbf{x}}$ and $N_{\mathbf{y}}$ are ad-hoc parameters (e.g. given an acoustic vector, why should there be precisely $N_{\mathbf{y}}$ candidates?).
- Training the MLP ensemble is difficult. Rahim et al. try several heuristic approaches to determine which MLP to adjust for a given speech frame.

Other methods have been recently proposed. In the analysis-by-synthesis technique of McGowan (1994) the articulatory model is the ASY articulatory synthesiser from the Haskins Laboratories (Rubin et al., 1981), which is based on the articulatory model of Mermelstein (1973). The acoustics are represented by the first three formants. A *task dynamics model* (Saltzman and Kelso, 1987) is added to further constrain the articulatory model: rather than articulatory trajectories, the task dynamics model uses "tract variables," which are variables describing constriction locations and degrees (that are more relevant than other parts of the vocal tract for determining the acoustics). The complexity of the task dynamics model makes difficult the use of derivative-based optimisation methods such as gradient descent. Thus, optimisation is performed by a genetic algorithm (Goldberg, 1989) with fitness = 1/error, where the error is the sum of squared errors for the first three formants. A disadvantage of genetic algorithms is that the parameters for optimisation must be coded into finite length strings (*chromosomes*) and this discretises the parameter space (to 6 bits per variable in his case). Using simulated data for /əbæ/ and /ədæ/ the method recovers most parts of an original articulatory trajectory but has trouble in obtaining precise timing, which is imputed to the lack of additional acoustic information, such as the RMS amplitude.

The approach of Sorokin and collaborators (see Sorokin, 1992; Sorokin et al., 2000 and references therein) is also an analysis-by-synthesis technique, where the inversion is considered from the point of view of standard regularisation theory (Tikhonov and Arsenin, 1977) for ill-posed problems (see chapter 6). In regularisation theory, the ill-posedness of an inverse problem is broken by the use of constraints. Sorokin et al. (2000) propose several types of constraints, most of which have been implicitly or explicitly used earlier in the acoustic-to-articulatory mapping problem; these include bounds on muscle forces, articulatory parameters and area functions, mutual dependence of the articulatory parameters (i.e., low intrinsic dimensionality) or complexity of planning and programming motor commands. The Tikhonov regularisation framework results in a cost function of the form acoustic fitness plus articulatory constraints, just as in eq. (10.2), and so the approach does not differ much from dynamic programming search of articulatory codebooks. Sorokin et al. have proposed articulatory models for vowels and fricatives that take into account a condition of non-turbulent air flow, given by a threshold for the Reynolds number, and minimise muscle effort as in eq. (10.3); and specific optimisation algorithms, since they use the uniform (L_∞) norm as acoustic distance between the formants, which is not differentiable.

Yehia and Itakura (1996) consider a simplified version of the acoustic-to-articulatory mapping problem where the acoustics are given by the first M formant frequencies and the articulator configurations by the first $2M$ coefficients of the Fourier cosine series expansion of the log-area function, for a known vocal tract length. Work by Mermelstein (1967) and Schroeder (1967) showed that a one-to-one relationship holds approximately between the first M odd Fourier coefficients of the area function and the first M formants and set to zero the M even Fourier coefficients, which are undetermined. Yehia and Itakura apply a quadratic cost function to the area function (representing minimal effort, like Sorokin, 1992) to obtain those M even coefficients; this is done at each frame, using the Newton-Raphson method, and no continuity constraint is used. Their results using a small corpus of French vowels are just barely better than those of Mermelstein (1967). This is not surprising since this approach simply converts the one-to-many mapping into a one-to-one mapping without even the guarantee that a solution branch is selected.

Based on the idea that speech production realises articulatory targets while subject to coarticulation, Blackburn and Young (2000) propose a method to produce a smooth articulatory trajectory given a time-aligned phonetic string. The trajectory is obtained from the requirement that it passes through *soft* regions of articulatory space given by each phoneme but keeping articulatory effort low. The soft regions are obtained by replacing the distribution of articulatory positions at the midpoint of a given phoneme (obtained from X-ray data) with an independent Gaussian distribution for each articulator. This is a variation of earlier work by Keating, who used *hard* windows (i.e., uniform distributions rather than Gaussian). The requirement that articulatory effort be low is attained by allowing the articulatory trajectory to under- or overshoot the midpoints (articulatory targets), resulting in a trajectory that is a smoothed version of the polygonal line joining the midpoints. This is in effect a coarticulation model, similar in its goal to that of Bakis (1993) (see section 10.1.4). The RMS errors for the recovered articulators' positions were around 35% of their standard deviation on the average. The selection of soft regions can be seen as a conditional mean method and one would expect it to perform similarly to an MLP with errorbars. Thus, for multimodal distributions, known to exist for the articulatory-to-acoustic mapping, the mean of the Gaussian may lie in an incorrect articulatory value.

Summarising, state-of-the-art articulatory inversion in terms of accuracy is obtained by dynamic programming search of a large, carefully constructed articulatory codebook, at an enormous computational cost in storage and time. A carefully prepared ensemble of neural networks approaches codebook performance and is fast.

10.1.4 Speech recognition models that incorporate production information

Hidden Markov models and variants of them[9] are currently the unrivalled method for automatic speech recognition. Like neural networks, HMMs are complex generic statistical models that could be used for the description of many physical phenomena because the strong assumptions that they make can usually be overcome by hav-

[9]In particular, hybrid recognisers based on neural nets and HMMs, which share the advantages of the two frameworks and often deliver superior performance (Bourlard and Morgan, 1994).

ing a large number of parameters and of training data. But there is a limit to what models based on acoustic information alone can do. The performance of HMMs degrades dramatically when the speech style changes, the speaker changes or there is noise[10], all of which occur in spontaneous speech in natural environments. Several people (e.g. Rose et al., 1996, Deng et al., 1997 and Deng, 1998) have suggested the addition of e.g. linguistic or production information to acoustic models. In particular, the advantages of articulatory representations discussed earlier and the availability of articulatory data from X-ray and EMA measurements have recently led to several models that incorporate articulatory constraints (not necessarily approaching the articulatory inversion). We briefly review some here.

Zlokarnik (1995a) has provided empirical evidence that straightforward addition of articulatory information to an acoustic HMM improves recognition performance. Simultaneously recorded acoustic and articulatory data (the positions of several articulators, recorded by EMA) were combined to make up an acoustic-articulatory feature vector on a speaker-dependent isolated word recognition task with an HMM. Compared with a purely acoustic HMM, using acoustic and articulatory data both for training and testing reduced the error rate by 60%; and using articulatory measurements only during the training and implementing an acoustic-to-articulatory mapping with an MLP during the testing phase, the error rate could be reduced by a relative percentage of 18% to 25%. In another experiment, Zlokarnik (1995b) showed that of the first three time derivatives (velocities, accelerations and jerks) of the articulators' positions, accelerations perform best for ASR. Although this is surprising, since in the acoustic domain, acceleration features perform worse than static features, it confirms the importance of the role of articulatory forces in speech production.

Other variations of HMMs that use articulatory information in more sophisticated ways (e.g. forbidding transitions) were discussed in section 7.11.5. But, given the continuous nature of the temporal variation of the articulators, it seems more natural to use models of the style of Kalman filters rather than HMMs.

Bakis (1993) (see also Bakis, 1991) proposed a generic speech production model that can be seen as an acoustic recognition model, such as an HMM, augmented by an analysis-by-synthesis technique. Given an acoustic waveform to be recognised, the acoustic model proposes a hypothesis, i.e., a phoneme transcription. An abstract, deterministic articulatory model then takes as input this transcription and synthesises acoustic features that can be compared with similar features computed from the actual speech. The abstract articulatory model works as follows. First, the phonetic string is transformed into an *idealised target path* in a multidimensional Euclidean space via a table lookup; this path is piecewise constant, with abrupt transitions at phoneme boundaries. Then, this path is transformed into a *realised articulatory path* in the same multidimensional Euclidean space via convolution with a FIR filter; this path is a smoothed version of the target path and results in bounded first and second derivatives of the articulators' motion (and in correspondingly bounded forces), thus modelling coarticulation. Finally, acoustic vectors are generated from the realised articulatory path via a neural network in the form of MFCCs or any other suitable acoustic representation. Therefore, the details of the vocal tract model are left to be determined empirically from data and only the general properties are specified: coarticulation is implemented by an empirical FIR filter (with memory) rather than described in terms of masses, forces and viscous damping; and the articulatory-to-acoustic mapping is implemented by an empirical neural network (memoryless nonlinear function) rather than derived from Webster's horn equation. The components of both the acoustic model (HMM) and the articulatory model (lookup table, filter, neural net) are parametric. The parameters are adjusted from prior knowledge and empirical information to minimise the mean square error of the acoustic vectors (using conjugate gradients, the gradient being computed by the chain rule). Further prior knowledge can be included as penalty terms on the parameters in the objective function. The time-aligned phonetic transcription is given as input at training time, while at recognition time it is proposed as a hypothesis to be tested.

In this model, then, the abstract space consists of a finite collection of targets—basically, an articulatory codebook—that acts as a scaffolding on which to create smooth trajectories. An important problem is thus how to select the dimensionality of the articulatory space and the number of phonetic targets, which must be given by the user. Presumably the number of phonetic targets is related to the number of different phonemes in the language or training set under consideration, but it needs not be necessarily equal (e.g. consider the case of diphthongs and allophones). Determining the dimensionality of the articulatory space is probably a similar problem to that of determining the map dimensionality in multidimensional scaling (section 4.10.1.1).

Bakis seems never to have implemented this interesting model in practice. Recently, Richards and Bridle (1999) have implemented essentially the same model, with two minor variations: the abstract articulatory model (which they rename *hidden dynamic model*) and the acoustic model are not trained jointly, but become

[10]In section 7.10.6 we described some research on occluded speech recognition based purely on acoustic models. Also, several techniques have been devised in the speech recognition literature that partially alleviate the problem of speaker adaptation, e.g. by maximum a-posteriori estimation (Gauvain and Lee, 1994) or maximum likelihood linear regression (Leggeter and Woodland, 1995).

separate entities, with the articulatory model being used to rescore N-best lists of hypotheses; and the realised articulatory path is obtained via a second-order symmetrical (forward-backward) low-pass filter with one time constant per articulatory dimension and phonetic target. This filter, which is a simple form of Kalman smoother, controls how much to undershoot a target: the larger the time constant, the more undershooting and smoothing; in the zero limit, the transitions are discontinuous and there is no smoothing. The filter is symmetrical so that the centre of transitions occurs at phone boundaries. Thus, the articulatory model remains deterministic and does not deal with time alignment (unknown time scales in phone durations). Richards and Bridle also show the necessity of a nonlinear articulatory-to-acoustic mapping: if a linear one is used instead, the model fails to produce continuous transitions. In an evaluation of the approach with a conversational speech recognition task with the Switchboard corpus (Picone et al., 1999), improvements in terms of word error rate compared to a standard acoustic HMM only occur if, as well as the most likely hypotheses, the reference transcription is given (which is unavailable in practice). This proves that the articulatory model has information not in the acoustic HMM, although it remains to be seen how to actually use it.

Deng (1998) has proposed a stochastic approach combining the two contrasting aspects of speech: phonological (characterised by the discrete nature of phonemes) and phonetic (characterised by the continuous nature of the vocal tract). The basic structure of the model is the same as that of Bakis (1993): the phonemic string is realised in a continuous, dynamical system and nonlinearly mapped onto the acoustic, observable features. Specifically, the model consists of the following levels: (1) a language model which provides the probability for an arbitrary word sequence $p(W = w_1 \cdots w_N)$; (2) a phonological or pronunciation model based, rather than on phones (as most speech recognisers are), on overlapping features (Browman and Goldstein, 1992), which provides the probability $p(F|W)$ for a feature-overlapping pattern F of an entire utterance given its word sequence; and (3) a phonetic model which provides the probability $p(O|F, W)$ of an observed acoustic trajectory O, based on the task dynamics model of Saltzman and Kelso (1987), which is implemented with a smooth linear dynamical system (with memory) and a nonlinear articulatory-to-acoustic mapping (memoryless, such as an MLP or RBF net). Consequently, inference about the word transcription W given the acoustics O is done via Bayes' rule as in HMMs. But here the dependence of the acoustic sequence on the word sequence is more complex, involving the intermediate stage of continuous variables at the phonetic level where the articulatory constraints are applied. This complex stochastic model, containing nonlinear functions and dynamical systems, seems to be trainable for maximum likelihood given observed acoustic data by a generalised EM algorithm—a surprising fact in view of the intractability that is invariably associated with the marginalisation of complex distributions. An evaluation of a version of this model with the mentioned Switchboard corpus (Picone et al., 1999) gave very similar results to those of the hidden dynamic model implemented by Richards and Bridle (1999).

King and Wrench (1999) and Frankel et al. (2000) have modelled the articulatory trajectories with a linear dynamical model (of 4 to 13 dimensions for the hidden space) and implemented the inversion mapping with a neural network similar to that of Papcun et al. (1992) but with the addition of recurrence via context units in a hidden layer, which results in smoother trajectories. They have used TIMIT sentences with acoustic and EMA data from the MOCHA database (section 7.10.5) in a recognition task. The results using acoustics plus recovered articulators' positions were considerably worse than using acoustics plus the real articulatory data or just using acoustic HMMs. One possible reason they adduce for this is that the segmentation based on acoustic information (data forced-aligned with an HMM, which assumes that state and phone boundaries are strictly synchronised with articulatory events) differs from the segmentation based on articulator positions: they observed a slight asynchrony between changes in articulatory gestures and HMM-produced phone boundaries.

In summary, while the articulatory trajectories contain information that can be used to improve automatic speech recognition, the integration of production and acoustic models has so far not attained this goal.

10.2 Experiments with electropalatographic and acoustic data

At the time when this research was being carried out, we did not have access to articulatory data that appropriately represented the vocal tract, either synthetic (from an articulatory model) or measured (with X-ray or EMA). Instead, we used the electropalatographic (EPG) data from the ACCOR database, as in chapter 5, together with the simultaneously recorded acoustic waveform. The EPG characterises well the pattern of tongue-palate contact but is an incomplete representation of the vocal tract, and so many phonemes are indistinguishable in the EPG. For example, in fig. 5.6, the EPG labelled /æ/ can result from many other vowels (and even from silence intervals), while those of /g/ and /k/ or /t/ and /d/ are almost interchangeable. Conversely, from the nonuniqueness of the acoustic-to-articulatory mapping it is also reasonable to assume that in certain cases one phoneme may be produced with more than one different EPG. Consequently, our

Bibliography

S. Amari. Natural gradient learning for over- and under-complete bases in ICA. *Neural Computation*, 11(8): 1875–1883, Nov. 1999.

S. Amari and A. Cichoki. Adaptive blind signal processing—neural network approaches. *Proc. IEEE*, 86(10): 2026–2048, Oct. 1998.

T. W. Anderson. Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, 34 (1):122–148, Mar. 1963.

T. W. Anderson and H. Rubin. Statistical inference in factor analysis. In J. Neyman, editor, *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability*, volume V, pages 111–150, Berkeley, 1956. University of California Press.

S. Arnfield. Artificial EPG palate image. The Reading EPG, 1995. Available online at `http://www.linguistics.reading.ac.uk/research/speechlab/epg/palate.jpg`, Feb. 1, 2000.

H. Asada and J.-J. E. Slotine. *Robot Analysis and Control*. John Wiley & Sons, New York, London, Sydney, 1986.

D. Asimov. The grand tour: A tool for viewing multidimensional data. *SIAM J. Sci. Stat. Comput.*, 6:128–143, 1985.

B. S. Atal, J. J. Chang, M. V. Mathews, and J. W. Tukey. Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *J. Acoustic Soc. Amer.*, 63(5):1535–1555, May 1978.

C. G. Atkeson. Learning arm kinematics and dynamics. *Annu. Rev. Neurosci.*, 12:157–183, 1989.

H. Attias. EM algorithms for independent component analysis. In Niranjan (1998), pages 132–141.

H. Attias. Independent factor analysis. *Neural Computation*, 11(4):803–851, May 1999.

F. Aurenhammer. Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Computing Surveys*, 23(3):345–405, Sept. 1991.

A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, B*, 61(3):579–602, 1999.

R. J. Baddeley. Searching for filters with "interesting" output distributions: An uninteresting direction to explore? *Network: Computation in Neural Systems*, 7(2):409–421, 1996.

R. Bakis. Coarticulation modeling with continuous-state HMMs. In *Proc. IEEE Workshop Automatic Speech Recognition*, pages 20–21, Arden House, New York, 1991. Harriman.

R. Bakis. An articulatory-like speech production model with controlled use of prior knowledge. Frontiers in Speech Processing: Robust Speech Analysis '93, Workshop CDROM, NIST Speech Disc 15 (also available from the Linguistic Data Consortium), Aug. 6 1993.

P. Baldi and K. Hornik. Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks*, 2(1):53–58, 1989.

J. D. Banfield and A. E. Raftery. Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *J. Amer. Stat. Assoc.*, 87(417):7–16, Mar. 1992.

J. Barker, L. Josifovski, M. Cooke, and P. Green. Soft decisions in missing data techniques for robust automatic speech recognition. In *Proc. of the International Conference on Spoken Language Processing (ICSLP'00)*, Beijing, China, Oct. 16–20 2000.

J. P. Barker and F. Berthommier. Evidence of correlation between acoustic and visual features of speech. In Ohala et al. (1999), pages 199–202.

M. F. Barnsley. *Fractals Everywhere*. Academic Press, New York, 1988.

D. J. Bartholomew. The foundations of factor analysis. *Biometrika*, 71(2):221–232, Aug. 1984.

D. J. Bartholomew. Foundations of factor analysis: Some practical implications. *Brit. J. of Mathematical and Statistical Psychology*, 38:1–10 (discussion in pp. 127–140), 1985.

D. J. Bartholomew. *Latent Variable Models and Factor Analysis.* Charles Griffin & Company Ltd., London, 1987.

A. Basilevsky. *Statistical Factor Analysis and Related Methods.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1994.

H.-U. Bauer, M. Herrmann, and T. Villmann. Neural maps and topographic vector quantization. *Neural Networks*, 12(4–5):659–676, June 1999.

H.-U. Bauer and K. R. Pawelzik. Quantifying the neighbourhood preservation of self-organizing feature maps. *IEEE Trans. Neural Networks*, 3(4):570–579, July 1992.

J. Behboodian. On the modes of a mixture of two normal distributions. *Technometrics*, 12(1):131–139, Feb. 1970.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

A. J. Bell and T. J. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res.*, 37(23):3327–3338, Dec. 1997.

R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, 1957.

R. Bellman. *Adaptive Control Processes: A Guided Tour.* Princeton University Press, Princeton, 1961.

Y. Bengio and F. Gingras. Recurrent neural networks for missing or asynchronous data. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 395–401. MIT Press, Cambridge, MA, 1996.

C. Benoît, M.-T. Lallouache, T. Mohamadi, and C. Abry. A set of French visemes for visual speech synthesis. In G. Bailly and C. Benoît, editors, *Talking Machines: Theories, Models and Designs*, pages 485–504. North Holland-Elsevier Science Publishers, Amsterdam, New York, Oxford, 1992.

P. M. Bentler and J. S. Tanaka. Problems with EM algorithms for ML factor analysis. *Psychometrika*, 48(2): 247–251, June 1983.

J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer Series in Statistics. Springer-Verlag, Berlin, second edition, 1985.

M. Berkane, editor. *Latent Variable Modeling and Applications to Causality.* Number 120 in Springer Series in Statistics. Springer-Verlag, Berlin, 1997.

J. M. Bernardo and A. F. M. Smith. *Bayesian Theory.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Chichester, 1994.

N. Bernstein. *The Coordination and Regulation of Movements.* Pergamon, Oxford, 1967.

D. P. Bertsekas. *Dynamic Programming. Deterministic and Stochastic Models.* Prentice-Hall, Englewood Cliffs, N.J., 1987.

J. Besag and P. J. Green. Spatial statistics and Bayesian computation. *Journal of the Royal Statistical Society, B*, 55(1):25–37, 1993.

J. C. Bezdek and N. R. Pal. An index of topological preservation for feature extraction. *Pattern Recognition*, 28(3):381–391, Mar. 1995.

E. L. Bienenstock, L. N. Cooper, and P. W. Munro. Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, 2(1):32–48, Jan. 1982.

C. M. Bishop. Mixture density networks. Technical Report NCRG/94/004, Neural Computing Research Group, Aston University, Feb. 1994. Available online at `http://www.ncrg.aston.ac.uk/Papers/postscript/NCRG_94_004.ps.Z`.

C. M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, New York, Oxford, 1995.

C. M. Bishop. Bayesian PCA. In Kearns et al. (1999), pages 382–388.

C. M. Bishop, G. E. Hinton, and I. G. D. Strachan. GTM through time. In *IEE Fifth International Conference on Artificial Neural Networks*, pages 111–116, 1997a.

C. M. Bishop and I. T. Nabney. Modeling conditional probability distributions for periodic variables. *Neural Computation*, 8(5):1123–1133, July 1996.

C. M. Bishop, M. Svensén, and C. K. I. Williams. Magnification factors for the SOM and GTM algorithms. In *WSOM'97: Workshop on Self-Organizing Maps*, pages 333–338, Finland, June 4–6 1997b. Helsinki University of Technology.

C. M. Bishop, M. Svensén, and C. K. I. Williams. Developments of the generative topographic mapping. *Neurocomputing*, 21(1–3):203–224, Nov. 1998a.

C. M. Bishop, M. Svensén, and C. K. I. Williams. GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234, Jan. 1998b.

C. M. Bishop and M. E. Tipping. A hierarchical latent variable model for data visualization. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 20(3):281–293, Mar. 1998.

A. Bjerhammar. *Theory of Errors and Generalized Matrix Inverses.* North Holland-Elsevier Science Publishers, Amsterdam, New York, Oxford, 1973.

C. S. Blackburn and S. Young. A self-learning predictive model of articulator movements during speech production. *J. Acoustic Soc. Amer.*, 107(3):1659–1670, Mar. 2000.

T. L. Boullion and P. L. Odell. *Generalized Inverse Matrices.* John Wiley & Sons, New York, London, Sydney, 1971.

H. Bourlard and Y. Kamp. Autoassociation by the multilayer perceptrons and singular value decomposition. *Biol. Cybern.*, 59(4–5):291–294, 1988.

H. Bourlard and N. Morgan. *Connectionist Speech Recognition. A Hybrid Approach.* Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1994.

M. Brand. Structure learning in conditional probability models via an entropic prior and parameter extinction. *Neural Computation*, 11(5):1155–1182, July 1999.

C. Bregler and S. M. Omohundro. Surface learning with applications to lip-reading. In Cowan et al. (1994), pages 43–50.

C. Bregler and S. M. Omohundro. Nonlinear image interpolation using manifold learning. In Tesauro et al. (1995), pages 973–980.

L. J. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, Aug. 1996.

S. P. Brooks. Markov chain Monte Carlo method and its application. *The Statistician*, 47(1):69–100, 1998.

C. P. Browman and L. M. Goldstein. Articulatory phonology: An overview. *Phonetica*, 49(3–4):155–180, 1992.

E. N. Brown, L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson. A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells. *J. Neurosci.*, 18(18):7411–7425, Sept. 1998.

G. J. Brown and M. Cooke. Computational auditory scene analysis. *Computer Speech and Language*, 8(4): 297–336, Oct. 1994.

D. Byrd, E. Flemming, C. A. Mueller, and C. C. Tan. Using regions and indices in EPG data reduction. *Journal of Speech and Hearing Research*, 38(4):821–827, Aug. 1995.

J.-F. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Letters on Signal Processing*, 4(4):112–114, Apr. 1997.

J.-F. Cardoso. Blind signal separation: Statistical principles. *Proc. IEEE*, 86(10):2009–2025, Oct. 1998.

M. Á. Carreira-Perpiñán. A review of dimension reduction techniques. Technical Report CS–96–09, Dept. of Computer Science, University of Sheffield, UK, Dec. 1996. Available online at `http://www.dcs.shef.ac.uk/~miguel/papers/cs-96-09.html`.

M. Á. Carreira-Perpiñán. Density networks for dimension reduction of continuous data: Analytical solutions. Technical Report CS–97–09, Dept. of Computer Science, University of Sheffield, UK, Apr. 1997. Available online at `http://www.dcs.shef.ac.uk/~miguel/papers/cs-97-09.html`.

M. Á. Carreira-Perpiñán. Mode-finding for mixtures of Gaussian distributions. Technical Report CS–99–03, Dept. of Computer Science, University of Sheffield, UK, Mar. 1999a. Revised August 4, 2000. Available online at `http://www.dcs.shef.ac.uk/~miguel/papers/cs-99-03.html`.

M. Á. Carreira-Perpiñán. One-to-many mappings, continuity constraints and latent variable models. In *Proc. of the IEE Colloquium on Applied Statistical Pattern Recognition*, Birmingham, UK, 1999b.

M. Á. Carreira-Perpiñán. Mode-finding for mixtures of Gaussian distributions. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 22(11):1318–1323, Nov. 2000a.

M. Á. Carreira-Perpiñán. Reconstruction of sequential data with probabilistic models and continuity constraints. In Solla et al. (2000), pages 414–420.

M. Á. Carreira-Perpiñán and S. Renals. Dimensionality reduction of electropalatographic data using latent variable models. *Speech Communication*, 26(4):259–282, Dec. 1998a.

M. Á. Carreira-Perpiñán and S. Renals. Experimental evaluation of latent variable models for dimensionality reduction. In Niranjan (1998), pages 165–173.

M. Á. Carreira-Perpiñán and S. Renals. A latent variable modelling approach to the acoustic-to-articulatory mapping problem. In Ohala et al. (1999), pages 2013–2016.

M. Á. Carreira-Perpiñán and S. Renals. Practical identifiability of finite mixtures of multivariate Bernoulli distributions. *Neural Computation*, 12(1):141–152, Jan. 2000.

J. Casti. Flight over Wall St. *New Scientist*, 154(2078):38–41, Apr. 19 1997.

T. Chen and R. R. Rao. Audio-visual integration in multimodal communication. *Proc. IEEE*, 86(5):837–852, May 1998.

H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *J. Amer. Stat. Assoc.*, 68(342):361–368, June 1973.

D. A. Cohn. Neural network exploration using optimal experiment design. *Neural Networks*, 9(6):1071–1083, Aug. 1996.

C. H. Coker. A model of articulatory dynamics and control. *Proc. IEEE*, 64(4):452–460, 1976.

P. Comon. Independent component analysis: A new concept? *Signal Processing*, 36(3):287–314, Apr. 1994.

S. C. Constable, R. L. Parker, and C. G. Constable. Occam's inversion—a practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52(3):289–300, 1987.

D. Cook, A. Buja, and J. Cabrera. Projection pursuit indexes based on orthonormal function expansions. *Journal of Computational and Graphical Statistics*, 2(3):225–250, 1993.

M. Cooke and D. P. W. Ellis. The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, 2000. To appear.

M. Cooke, P. Green, L. Josifovski, and A. Vizinho. Robust automatic speech recognition with missing and unreliable acoustic data. *Speech Communication*, 34(3):267–285, June 2001.

D. Cornford, I. T. Nabney, and D. J. Evans. Bayesian retrieval of scatterometer wind fields. Technical Report NCRG/99/015, Neural Computing Research Group, Aston University, 1999a. Submitted to J. of Geophysical Research. Available online at `ftp://cs.aston.ac.uk/cornford/bayesret.ps.gz`.

D. Cornford, I. T. Nabney, and C. K. I. Williams. Modelling frontal discontinuities in wind fields. Technical Report NCRG/99/001, Neural Computing Research Group, Aston University, Jan. 1999b. Submitted to Nonparametric Statistics. Available online at `http://www.ncrg.aston.ac.uk/Papers/postscript/NCRG_99_001.ps.Z`.

R. Courant and D. Hilbert. *Methods of Mathematical Physics*, volume 1. Interscience Publishers, New York, 1953.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, New York, London, Sydney, 1991.

J. D. Cowan, G. Tesauro, and J. Alspector, editors. *Advances in Neural Information Processing Systems*, volume 6, 1994. Morgan Kaufmann, San Mateo.

T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman & Hall, London, New York, 1994.

J. J. Craig. *Introduction to Robotics. Mechanics and Control*. Series in Electrical and Computer Engineering: Control Engineering. Addison-Wesley, Reading, MA, USA, second edition, 1989.

N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, 2000.

P. Dayan. Arbitrary elastic topologies and ocular dominance. *Neural Computation*, 5(3):392–401, 1993.

P. Dayan, G. E. Hinton, R. M. Neal, and R. S. Zemel. The Helmholtz machine. *Neural Computation*, 7(5): 889–904, Sept. 1995.

M. H. DeGroot. *Probability and Statistics*. Addison-Wesley, Reading, MA, USA, 1986.

D. DeMers and G. W. Cottrell. Non-linear dimensionality reduction. In S. J. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances in Neural Information Processing Systems*, volume 5, pages 580–587. Morgan Kaufmann, San Mateo, 1993.

D. DeMers and K. Kreutz-Delgado. Learning global direct inverse kinematics. In J. Moody, S. J. Hanson, and R. P. Lippmann, editors, *Advances in Neural Information Processing Systems*, volume 4, pages 589–595. Morgan Kaufmann, San Mateo, 1992.

D. DeMers and K. Kreutz-Delgado. Canonical parameterization of excess motor degrees of freedom with self-organizing maps. *IEEE Trans. Neural Networks*, 7(1):43–55, Jan. 1996.

D. DeMers and K. Kreutz-Delgado. Learning global properties of nonredundant kinematic mappings. *Int. J. of Robotics Research*, 17(5):547–560, May 1998.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the *EM* algorithm. *Journal of the Royal Statistical Society, B*, 39(1):1–38, 1977.

L. Deng. A dynamic, feature-based approach to the interface between phonology and phonetics for speech modeling and recognition. *Speech Communication*, 24(4):299–323, July 1998.

L. Deng, G. Ramsay, and D. Sun. Production models as a structural basis for automatic speech recognition. *Speech Communication*, 22(2–3):93–111, Aug. 1997.

P. Diaconis and D. Freedman. Asymptotics of graphical projection pursuit. *Annals of Statistics*, 12(3):793–815, Sept. 1984.

K. I. Diamantaras and S.-Y. Kung. *Principal Component Neural Networks. Theory and Applications*. Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, New York, London, Sydney, 1996.

T. G. Dietterich. Machine learning research: Four current directions. *AI Magazine*, 18(4):97–136, winter 1997.

M. P. do Carmo. *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Englewood Cliffs, N.J., 1976.

R. D. Dony and S. Haykin. Optimally adaptive transform coding. *IEEE Trans. on Image Processing*, 4(10): 1358–1370, Oct. 1995.

R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley & Sons, New York, London, Sydney, 1973.

R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, editors. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

R. Durbin and G. Mitchison. A dimension reduction framework for understanding cortical maps. *Nature*, 343 (6259):644–647, Feb. 15 1990.

R. Durbin, R. Szeliski, and A. Yuille. An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation*, 1(3):348–358, Fall 1989.

R. Durbin and D. Willshaw. An analogue approach to the traveling salesman problem using an elastic net method. *Nature*, 326(6114):689–691, Apr. 16 1987.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers Group, Dordrecht, The Netherlands, 1996.

K. Erler and G. H. Freeman. An HMM-based speech recognizer using overlapping articulatory features. *J. Acoustic Soc. Amer.*, 100(4):2500–2513, Oct. 1996.

G. Eslava and F. H. C. Marriott. Some criteria for projection pursuit. *Statistics and Computing*, 4:13–20, 1994.

C. Y. Espy-Wilson, S. E. Boyce, M. Jackson, S. Narayanan, and A. Alwan. Acoustic modeling of American English /r/. *J. Acoustic Soc. Amer.*, 108(1):343–356, July 2000.

J. Etezadi-Amoli and R. P. McDonald. A second generation nonlinear factor analysis. *Psychometrika*, 48(3): 315–342, Sept. 1983.

D. J. Evans, D. Cornford, and I. T. Nabney. Structured neural network modelling of multi-valued functions for wind vector retrieval from satellite scatterometer measurements. *Neurocomputing*, 30(1–4):23–30, Jan. 2000.

B. S. Everitt. *An Introduction to Latent Variable Models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, New York, 1984.

B. S. Everitt and D. J. Hand. *Finite Mixture Distributions*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, New York, 1981.

K. J. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons, Chichester, 1990.

K. Fan. On a theorem of Weyl concerning the eigenvalues of linear transformations II. *Proc. Natl. Acad. Sci. USA*, 36:31–35, 1950.

G. Fant. *Acoustic Theory of Speech Production*. Mouton, The Hague, Paris, second edition, 1970.

E. Farnetani, W. J. Hardcastle, and A. Marchal. Cross-language investigation of lingual coarticulatory processes using EPG. In J.-P. Tubach and J.-J. Mariani, editors, *Proc. EUROSPEECH'89*, volume 2, pages 429–432, Paris, France, Sept. 26–28 1989.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2 of *Wiley Series in Probability and Mathematical Statistics*. John Wiley & Sons, New York, London, Sydney, third edition, 1971.

D. J. Field. What is the goal of sensory coding? *Neural Computation*, 6(4):559–601, July 1994.

J. L. Flanagan. *Speech Analysis, Synthesis and Perception*. Number 3 in Kommunication und Kybernetik in Einzeldarstellungen. Springer-Verlag, Berlin, second edition, 1972.

M. K. Fleming and G. W. Cottrell. Categorization of faces using unsupervised feature extraction. In *Proc. Int. J. Conf. on Neural Networks (IJCNN90)*, volume II, pages 65–70, San Diego, CA, June 17–21 1990.

P. Földiák. Adaptive network for optimal linear feature extraction. In *Proc. Int. J. Conf. on Neural Networks (IJCNN89)*, volume I, pages 401–405, Washington, DC, June 18–22 1989.

D. Fotheringhame and R. Baddeley. Nonlinear principal components analysis of neuronal data. *Biol. Cybern.*, 77(4):283–288, 1997.

I. E. Frank and J. H. Friedman. A statistical view of some chemometrics regression tools. *Technometrics*, 35 (2):109–135 (with comments: pp. 136–148), May 1993.

J. Frankel, K. Richmond, S. King, and P. Taylor. An automatic speech recognition system using neural networks and linear dynamic models to recover and model articulatory traces. In *Proc. of the International Conference on Spoken Language Processing (ICSLP'00)*, Beijing, China, Oct. 16–20 2000.

J. H. Friedman. Exploratory projection pursuit. *J. Amer. Stat. Assoc.*, 82(397):249–266, Mar. 1987.

J. H. Friedman. Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–67 (with comments, pp. 67–141), Mar. 1991.

J. H. Friedman and N. I. Fisher. Bump hunting in high-dimensional data. *Statistics and Computing*, 9(2): 123–143 (with discussion, pp. 143–162), Apr. 1999.

J. H. Friedman and W. Stuetzle. Projection pursuit regression. *J. Amer. Stat. Assoc.*, 76(376):817–823, Dec. 1981.

J. H. Friedman, W. Stuetzle, and A. Schroeder. Projection pursuit density estimation. *J. Amer. Stat. Assoc.*, 79(387):599–608, Sept. 1984.

J. H. Friedman and J. W. Tukey. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Computers*, C–23:881–889, 1974.

C. Fyfe and R. J. Baddeley. Finding compact and sparse distributed representations of visual images. *Network: Computation in Neural Systems*, 6(3):333–344, Aug. 1995.

J.-L. Gauvain and C.-H. Lee. Maximum a-posteriori estimation for multivariate Gaussian mixture observations of Markov chains. *IEEE Trans. Speech and Audio Process.*, 2:1291–1298, 1994.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Texts in Statistical Science. Chapman & Hall, London, New York, 1995.

C. Genest and J. V. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–135 (with discussion, pp. 135–148), Feb. 1986.

Z. Ghahramani. Solving inverse problems using an EM approach to density estimation. In M. C. Mozer, P. Smolensky, D. S. Touretzky, J. L. Elman, and A. S. Weigend, editors, *Proceedings of the 1993 Connectionist Models Summer School*, pages 316–323, 1994.

Z. Ghahramani and M. J. Beal. Variational inference for Bayesian mixture of factor analysers. In Solla et al. (2000), pages 449–455.

Z. Ghahramani and G. E. Hinton. The EM algorithm for mixtures of factor analyzers. Technical Report CRG–TR–96–1, University of Toronto, May 21 1996. Available online at `ftp://ftp.cs.toronto.edu/pub/zoubin/tr-96-1.ps.gz`.

Z. Ghahramani and M. I. Jordan. Supervised learning from incomplete data via an EM approach. In Cowan et al. (1994), pages 120–127.

W. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London, New York, 1996.

M. Girolami, A. Cichoki, and S. Amari. A common neural network model for exploratory data analysis and independent component analysis. *IEEE Trans. Neural Networks*, 9(6):1495–1501, 1998.

M. Girolami and C. Fyfe. Stochastic ICA contrast maximization using Oja's nonlinear PCA algorithm. *Int. J. Neural Syst.*, 8(5–6):661–678, Oct./Dec. 1999.

F. Girosi, M. Jones, and T. Poggio. Regularization theory and neural networks architectures. *Neural Computation*, 7(2):219–269, Mar. 1995.

S. J. Godsill and P. J. W. Rayner. *Digital Audio Restoration: A Statistical Model-Based Approach*. Springer-Verlag, Berlin, 1998a.

S. J. Godsill and P. J. W. Rayner. Robust reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Trans. Speech and Audio Process.*, 6(4):352–372, July 1998b.

B. Gold and N. Morgan. *Speech and Audio Signal Processing: Processing and Perception of Speech and Music*. John Wiley & Sons, New York, London, Sydney, 2000.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.

G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins Press, Baltimore, third edition, 1996.

G. J. Goodhill and T. J. Sejnowski. A unifying objective function for topographic mappings. *Neural Computation*, 9(6):1291–1303, Aug. 1997.

R. A. Gopinath, B. Ramabhadran, and S. Dharanipragada. Factor analysis invariant to linear transformations of data. In *Proc. of the International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, Nov. 30 – Dec. 4 1998.

W. P. Gouveia and J. A. Scales. Resolution of seismic waveform inversion: Bayes versus Occam. *Inverse Problems*, 13(2):323–349, Apr. 1997.

W. P. Gouveia and J. A. Scales. Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *J. of Geophysical Research*, 130(B2):2759–2779, 1998.

I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, San Diego, fifth edition, 1994. Corrected and enlarged edition, edited by Alan Jeffrey.

R. M. Gray. Vector quantization. *IEEE ASSP Magazine*, pages 4–29, Apr. 1984.

R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. Inf. Theory*, 44(6):2325–2383, Oct. 1998.

M. Gyllenberg, T. Koski, E. Reilink, and M. Verlaan. Non-uniqueness in probabilistic numerical identification of bacteria. *J. Appl. Prob.*, 31:542–548, 1994.

P. Hall. On polynomial-based projection indices for exploratory projection pursuit. *Annals of Statistics*, 17 (2):589–605, June 1989.

W. J. Hardcastle, F. E. Gibbon, and W. Jones. Visual display of tongue-palate contact: Electropalatography in the assessment and remediation of speech disorders. *Brit. J. of Disorders of Communication*, 26:41–74, 1991a.

W. J. Hardcastle, F. E. Gibbon, and K. Nicolaidis. EPG data reduction methods and their implications for studies of lingual coarticulation. *J. of Phonetics*, 19:251–266, 1991b.

W. J. Hardcastle and N. Hewlett, editors. *Coarticulation: Theory, Data, and Techniques*. Cambridge Studies in Speech Science and Communication. Cambridge University Press, Cambridge, U.K., 1999.

W. J. Hardcastle, W. Jones, C. Knight, A. Trudgeon, and G. Calder. New developments in electropalatography: A state-of-the-art report. *J. Clinical Linguistics and Phonetics*, 3:1–38, 1989.

H. H. Harman. *Modern Factor Analysis*. University of Chicago Press, Chicago, second edition, 1967.

A. C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, 1991.

T. J. Hastie and W. Stuetzle. Principal curves. *J. Amer. Stat. Assoc.*, 84(406):502–516, June 1989.

T. J. Hastie and R. J. Tibshirani. *Generalized Additive Models.* Number 43 in Monographs on Statistics and Applied Probability. Chapman & Hall, London, New York, 1990.

G. T. Herman. *Image Reconstruction from Projections. The Fundamentals of Computer Tomography.* Academic Press, New York, 1980.

H. Hermansky. Perceptual linear predictive (PLP) analysis of speech. *J. Acoustic Soc. Amer.*, 87(4):1738–1752, Apr. 1990.

H. Hermansky and N. Morgan. RASTA processing of speech. *IEEE Trans. Speech and Audio Process.*, 2(4): 578–589, Oct. 1994.

J. A. Hertz, A. S. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation.* Number 1 in Santa Fe Institute Studies in the Sciences of Complexity Lecture Notes. Addison-Wesley, Reading, MA, USA, 1991.

G. E. Hinton. Products of experts. In D. Wilshaw, editor, *Proc. of the Ninth Int. Conf. on Artificial Neural Networks (ICANN99)*, pages 1–6, Edinburgh, UK, Sept. 7–10 1999. The Institution of Electrical Engineers.

G. E. Hinton, P. Dayan, and M. Revow. Modeling the manifolds of images of handwritten digits. *IEEE Trans. Neural Networks*, 8(1):65–74, Jan. 1997.

T. Holst, P. Warren, and F. Nolan. Categorising [s], [ʃ] and intermediate electropalographic patterns: Neural networks and other approaches. *European Journal of Disorders of Communication*, 30(2):161–174, 1995.

H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. of Educational Psychology*, 24:417–441 and 498–520, 1933.

P. J. Huber. *Robust Statistics.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1981.

P. J. Huber. Projection pursuit. *Annals of Statistics*, 13(2):435–475 (with comments, pp. 475–525), June 1985.

D. Husmeier. *Neural Networks for Conditional Probability Estimation.* Perspectives in Neural Computing. Springer-Verlag, Berlin, 1999.

J.-N. Hwang, S.-R. Lay, M. Maechler, R. D. Martin, and J. Schimert. Regression modeling in back-propagation and projection pursuit learning. *IEEE Trans. Neural Networks*, 5(3):342–353, May 1994.

A. Hyvärinen. New approximations of differential entropy for independent component analysis and projection pursuit. In Jordan et al. (1998), pages 273–279.

A. Hyvärinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks*, 10(3):626–634, Oct. 1999a.

A. Hyvärinen. Survey on independent component analysis. *Neural Computing Surveys*, 2:94–128, 1999b.

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis.* Wiley Series on Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, New York, London, Sydney, 2001.

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13(4–5):411–430, June 2000.

N. Intrator and L. N. Cooper. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5(1):3–17, 1992.

E. Isaacson and H. B. Keller. *Analysis of Numerical Methods.* John Wiley & Sons, New York, London, Sydney, 1966.

M. Isard and A. Blake. CONDENSATION — conditional density propagation for visual tracking. *Int. J. Computer Vision*, 29(1):5–28, 1998.

J. E. Jackson. *A User's Guide to Principal Components*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1991.

R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991.

M. Jamshidian and P. M. Bentler. A quasi-Newton method for minimum trace factor analysis. *J. of Statistical Computation and Simulation*, 62(1–2):73–89, 1998.

N. Japkowicz, S. J. Hanson, and M. A. Gluck. Nonlinear autoassociation is not equivalent to PCA. *Neural Computation*, 12(3):531–545, Mar. 2000.

E. T. Jaynes. Prior probabilities. *IEEE Trans. Systems, Science, and Cybernetics*, SSC-4(3):227–241, 1968.

F. V. Jensen. *An Introduction to Bayesian Networks*. UCL Press, London, 1996.

I. T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer-Verlag, Berlin, 1986.

M. C. Jones. *The Projection Pursuit Algorithm for Exploratory Data Analysis*. PhD thesis, University of Bath, 1983.

M. C. Jones and R. Sibson. What is projection pursuit? *Journal of the Royal Statistical Society, A*, 150(1): 1–18 (with comments, pp. 19–36), 1987.

W. Jones and W. J. Hardcastle. New developments in EPG3 software. *European Journal of Disorders of Communication*, 30(2):183–192, 1995.

M. I. Jordan. Motor learning and the degrees of freedom problem. In M. Jeannerod, editor, *Attention and Performance XIII*, pages 796–836. Lawrence Erlbaum Associates, Hillsdale, New Jersey and London, 1990.

M. I. Jordan, editor. *Learning in Graphical Models*, Adaptive Computation and Machine Learning series, 1998. MIT Press. Proceedings of the NATO Advanced Study Institute on Learning in Graphical Models, held in Erice, Italy, September 27 – October 7, 1996.

M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, and L. K. Saul. An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233, Nov. 1999.

M. I. Jordan and R. A. Jacobs. Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6(2):181–214, Mar. 1994.

M. I. Jordan, M. J. Kearns, and S. A. Solla, editors. *Advances in Neural Information Processing Systems*, volume 10, 1998. MIT Press, Cambridge, MA.

M. I. Jordan and D. E. Rumelhart. Forward models: Supervised learning with a distal teacher. *Cognitive Science*, 16(3):307–354, July–Sept. 1992.

K. G. Jöreskog. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4):443–482, Dec. 1967.

K. G. Jöreskog. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34 (2):183–202, June 1969.

H. F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, Sept. 1958.

N. Kambhatla and T. K. Leen. Dimension reduction by local principal component analysis. *Neural Computation*, 9(7):1493–1516, Oct. 1997.

G. K. Kanji. 100 *Statistical Tests*. Sage Publications, London, 1993.

J. N. Kapur. *Maximum-Entropy Models in Science and Engineering*. John Wiley & Sons, New York, London, Sydney, 1989.

J. Karhunen and J. Joutsensalo. Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, 7(1):113–127, 1994.

R. E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *J. Amer. Stat. Assoc.*, 91 (435):1343–1370, Sept. 1996.

M. S. Kearns, S. A. Solla, and D. A. Cohn, editors. *Advances in Neural Information Processing Systems*, volume 11, 1999. MIT Press, Cambridge, MA.

B. Kégl, A. Krzyzak, T. Linder, and K. Zeger. Learning and design of principal curves. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 22(3):281–297, Mar. 2000.

M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics Vol. 1: Distribution Theory.* Charles Griffin & Company Ltd., London, fourth edition, 1977.

W. M. Kier and K. K. Smith. Tongues, tentacles and trunks: The biomechanics of movement in muscular-hydrostats. *Zoological Journal of the Linnean Society*, 83:307–324, 1985.

S. King and A. Wrench. Dynamical system modelling of articulator movement. In Ohala et al. (1999), pages 2259–2262.

B. E. D. Kingsbury, N. Morgan, and S. Greenberg. Robust speech recognition using the modulation spectrogram. *Speech Communication*, 25(1–3):117–132, Aug. 1998.

T. K. Kohonen. *Self-Organizing Maps.* Number 30 in Springer Series in Information Sciences. Springer-Verlag, Berlin, 1995.

A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayner. Interpolation of missing data in image sequences. *IEEE Trans. on Image Processing*, 4(11):1509–1519, Nov. 1995.

J. F. Kolen and J. B. Pollack. Back propagation is sensitive to initial conditions. *Complex Systems*, 4(3): 269–280, 1990.

A. C. Konstantellos. Unimodality conditions for Gaussian sums. *IEEE Trans. Automat. Contr.*, AC–25(4): 838–839, Aug. 1980.

M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *Journal of the American Institute of Chemical Engineers*, 37(2):233–243, Feb. 1991.

J. B. Kruskal and M. Wish. *Multidimensional Scaling.* Number 07–011 in Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage Publications, Beverly Hills, 1978.

W. J. Krzanowski. *Principles of Multivariate Analysis: A User's Perspective.* Number 3 in Oxford Statistical Science Series. Oxford University Press, New York, Oxford, 1988.

S. Y. Kung, K. I. Diamantaras, and J. S. Taur. Adaptive principal component extraction (APEX) and applications. *IEEE Trans. Signal Processing*, 42(5):1202–1217, May 1994.

O. M. Kvalheim. The latent variable. *Chemometrics and Intelligent Laboratory Systems*, 14:1–3, 1992.

P. Ladefoged. Articulatory parameters. *Language and Speech*, 23(1):25–30, Jan.–Mar. 1980.

P. Ladefoged. *A Course in Phonetics.* Harcourt College Publishers, Fort Worth, fourth edition, 2000.

N. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *J. Amer. Stat. Assoc.*, 73 (364):805–811, Dec. 1978.

J. N. Larar, J. Schroeter, and M. M. Sondhi. Vector quantisation of the articulatory space. *IEEE Trans. Acoust., Speech, and Signal Process.*, ASSP-36(12):1812–1818, Dec. 1988.

F. Lavagetto. Time-delay neural networks for estimating lip movements from speech analysis: A useful tool in audio-video synchronization. *IEEE Trans. Circuits and Systems for video technology*, 7(5):786–800, Oct. 1997.

E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors. *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Chichester, England, 1985.

D. N. Lawley. A modified method of estimation in factor analysis and some large sample results. *Nord. Psykol. Monogr. Ser.*, 3:35–42, 1953.

P. F. Lazarsfeld and N. W. Henry. *Latent Structure Analysis*. Houghton-Mifflin, Boston, 1968.

M. LeBlanc and R. Tibshirani. Adaptive principal surfaces. *J. Amer. Stat. Assoc.*, 89(425):53–64, Mar. 1994.

D. D. Lee and H. Sompolinsky. Learning a continuous hidden variable model for binary data. In Kearns et al. (1999), pages 515–521.

T.-W. Lee, M. Girolami, and T. J. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, 11(2):417–441, Feb. 1999.

C. J. Leggeter and P. C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech and Language*, 9(2):171–185, Apr. 1995.

S. E. Levinson and C. E. Schmidt. Adaptive computation of articulatory parameters from the speech signal. *J. Acoustic Soc. Amer.*, 74(4):1145–1154, Oct. 1983.

M. S. Lewicki and T. J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, Feb. 2000.

A. M. Liberman, F. S. Cooper, D. P. Shankweiler, and M. Studdert-Kennedy. Perception of the speech code. *Psychological Review*, 74(6):431–461, 1967.

A. M. Liberman and I. G. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.

B. G. Lindsay. The geometry of mixture likelihoods: A general theory. *Annals of Statistics*, 11(1):86–94, Mar. 1983.

R. Linsker. An application of the principle of maximum information preservation to linear systems. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems*, volume 1, pages 186–194. Morgan Kaufmann, San Mateo, 1989.

R. J. A. Little. Regression with missing X's: A review. *J. Amer. Stat. Assoc.*, 87(420):1227–1237, Dec. 1992.

R. J. A. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1987.

S. P. Luttrell. A Bayesian analysis of self-organizing maps. *Neural Computation*, 6(5):767–794, Sept. 1994.

D. J. C. MacKay. Bayesian interpolation. *Neural Computation*, 4(3):415–447, May 1992a.

D. J. C. MacKay. A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3): 448–472, May 1992b.

D. J. C. MacKay. Bayesian neural networks and density networks. *Nuclear Instruments and Methods in Physics Research A*, 354(1):73–80, Jan. 1995a.

D. J. C. MacKay. Probable networks and plausible predictions — a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3):469–505, 1995b.

D. J. C. MacKay. Maximum likelihood and covariant algorithms for independent component analysis. Draft 3.7, Cavendish Laboratory, University of Cambridge, Dec. 19 1996. Available online at `http://wol.ra.phy.cam.ac.uk/mackay/abstracts/ica.html`.

D. J. C. MacKay. Comparison of approximate methods for handling hyperparameters. *Neural Computation*, 11(5):1035–1068, July 1999.

S. Maeda. A digital simulation method of the vocal tract system. *Speech Communication*, 1(3–4):199–229, 1982.

S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, and T. J. Sejnowski. Blind separation of auditory event-related brain responses into independent components. *Proc. Natl. Acad. Sci. USA*, 94:10979–10984, Sept. 1997.

E. C. Malthouse. Limitations of nonlinear PCA as performed with generic neural networks. *IEEE Trans. Neural Networks*, 9(1):165–173, Jan. 1998.

J. Mao and A. K. Jain. Artificial neural networks for feature extraction and multivariate data projection. *IEEE Trans. Neural Networks*, 6(2):296–317, Mar. 1995.

A. Marchal and W. J. Hardcastle. ACCOR: Instrumentation and database for the cross-language study of coarticulation. *Language and Speech*, 36(2, 3):137–153, 1993.

K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Probability and Mathematical Statistics Series. Academic Press, New York, 1979.

A. D. Marrs and A. R. Webb. Exploratory data analysis using radial basis function latent variable models. In Kearns et al. (1999), pages 529–535.

T. M. Martinetz and K. J. Schulten. Topology representing networks. *Neural Networks*, 7(3):507–522, 1994.

G. P. McCabe. Principal variables. *Technometrics*, 26(2):137–144, 1984.

R. P. McDonald. *Factor Analysis and Related Methods*. Lawrence Erlbaum Associates, Hillsdale, New Jersey and London, 1985.

R. S. McGowan. Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: Preliminary model tests. *Speech Communication*, 14(1):19–48, Feb. 1994.

R. S. McGowan and A. Faber. Introduction to papers on speech recognition and perception from an articulatory point of view. *J. Acoustic Soc. Amer.*, 99(3):1680–1682, Mar. 1996.

G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1997.

G. J. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 2000.

X.-L. Meng and D. van Dyk. The EM algorithm — an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society, B*, 59(3):511–540 (with discussion, pp. 541–567), 1997.

P. Mermelstein. Determination of vocal-tract shape from measured formant frequencies. *J. Acoustic Soc. Amer.*, 41(5):1283–1294, 1967.

P. Mermelstein. Articulatory model for the study of speech production. *J. Acoustic Soc. Amer.*, 53(4):1070–1082, 1973.

L. Mirsky. *An Introduction to Linear Algebra*. Clarendon Press, Oxford, 1955. Reprinted in 1982 by Dover Publications.

B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 19(7):696–710, July 1997.

J. Moody and C. J. Darken. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1(2):281–294, Summer 1989.

D. F. Morrison. *Multivariate Statistical Methods*. McGraw-Hill, New York, third edition, 1990.

K. Mosegaard and A. Tarantola. Monte-Carlo sampling of solutions to inverse problems. *J. of Geophysical Research—Solid Earth*, 100(B7):12431–12447, 1995.

É. Moulines, J.-F. Cardoso, and E. Gassiat. Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP'97)*, volume 5, pages 3617–3620, Munich, Germany, Apr. 21–24 1997.

J. R. Movellan, P. Mineiro, and R. J. Williams. Modeling path distributions using partially observable diffusion networks: A Monte-Carlo approach. Technical Report 99.01, Department of Cognitive Science, University of California, San Diego, June 1999. Available online at `http://hci.ucsd.edu/cogsci/tech_reports/faculty_pubs/99_01.ps`.

F. Mulier and V. Cherkassky. Self-organization as an iterative kernel smoothing process. *Neural Computation*, 7(6):1165–1177, Nov. 1995.

I. T. Nabney, D. Cornford, and C. K. I. Williams. Bayesian inference for wind field retrieval. *Neurocomputing*, 30(1–4):3–11, Jan. 2000.

J.-P. Nadal and N. Parga. Non-linear neurons in the low-noise limit: A factorial code maximizes information transfer. *Network: Computation in Neural Systems*, 5(4):565–581, Nov. 1994.

R. M. Neal. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG–TR–93–1, Dept. of Computer Science, University of Toronto, Sept. 1993. Available online at `ftp://ftp.cs.toronto.edu/pub/radford/review.ps.Z`.

R. M. Neal. *Bayesian Learning for Neural Networks*. Springer Series in Statistics. Springer-Verlag, Berlin, 1996.

R. M. Neal and P. Dayan. Factor analysis using delta-rule wake-sleep learning. *Neural Computation*, 9(8): 1781–1803, Nov. 1997.

R. M. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In Jordan (1998), pages 355–368. Proceedings of the NATO Advanced Study Institute on Learning in Graphical Models, held in Erice, Italy, September 27 – October 7, 1996.

W. L. Nelson. Physical principles for economies of skilled movements. *Biol. Cybern.*, 46(2):135–147, 1983.

N. Nguyen. EPG bidimensional data reduction. *European Journal of Disorders of Communication*, 30:175–182, 1995.

N. Nguyen, P. Hoole, and A. Marchal. Regenerating the spectral shape of [s] and [ʃ] from a limited set of articulatory parameters. *J. Acoustic Soc. Amer.*, 96(1):33–39, July 1994.

N. Nguyen, A. Marchal, and A. Content. Modeling tongue-palate contact patterns in the production of speech. *J. of Phonetics*, 24(1):77–97, Jan. 1996.

K. Nicolaidis and W. J. Hardcastle. Articulatory-acoustic analysis of selected English sentences from the EUR-ACCOR corpus. Technical report, SPHERE (Human capital and mobility program), 1994.

K. Nicolaidis, W. J. Hardcastle, A. Marchal, and N. Nguyen-Trong. Comparing phonetic, articulatory, acoustic and aerodynamic signal representations. In M. Cooke, S. Beet, and M. Crawford, editors, *Visual Representations of Speech Signals*, pages 55–82. John Wiley & Sons, 1993.

M. A. L. Nicolelis. Actions from thoughts. *Nature*, 409(6818):403–407, Jan. 18 2001.

M. Niranjan, editor. *Proc. of the 1998 IEEE Signal Processing Society Workshop on Neural Networks for Signal Processing (NNSP98)*, Cambridge, UK, Aug. 31 – Sept. 2 1998.

D. A. Nix and J. E. Hogden. Maximum-likelihood continuity mapping (MALCOM): An alternative to HMMs. In Kearns et al. (1999), pages 744–750.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, 1999.

J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. C. Bailey, editors. *Proc. of the 14th International Congress of Phonetic Sciences (ICPhS'99)*, San Francisco, USA, Aug. 1–7 1999.

E. Oja. Principal components, minor components, and linear neural networks. *Neural Networks*, 5(6):927–935, Nov.–Dec. 1992.

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, June 13 1996.

B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.*, 37(23):3311–3325, Dec. 1997.

M. W. Oram, P. Földiák, D. I. Perret, and F. Sengpiel. The 'ideal homunculus': Decoding neural population signals. *Trends Neurosci.*, 21(6):259–265, June 1998.

D. Ormoneit and V. Tresp. Penalized likelihood and Bayesian estimation for improving Gaussian mixture probability density estimates. *IEEE Trans. Neural Networks*, 9(4):639–650, July 1998.

M. Ostendorf, V. V. Digalakis, and O. A. Kimball. From HMM's to segment models: A unified view of stochastic modeling for speech recognition. *IEEE Trans. Speech and Audio Process.*, 4(5):360–378, Sept. 1996.

G. Papcun, J. Hochberg, T. R. Thomas, F. Laroche, J. Zacks, and S. Levy. Inferring articulation and recognizing gestures from acoustics with a neural network trained on x-ray microbeam data. *J. Acoustic Soc. Amer.*, 92(2):688–700, Aug. 1992.

J. Park and I. W. Sandberg. Approximation and radial-basis-function networks. *Neural Computation*, 5(2): 305–316, Mar. 1993.

R. L. Parker. *Geophysical Inverse Theory*. Princeton University Press, Princeton, 1994.

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, 1988.

B. A. Pearlmutter. Gradient calculation for dynamic recurrent neural networks: A survey. *IEEE Trans. Neural Networks*, 6(5):1212–1228, 1995.

B. A. Pearlmutter and L. C. Parra. A context-sensitive generalization of ICA. In *International Conference on Neural Information Processing (ICONIP–96), Hong Kong*, pages 151–157, Sept. 1996.

K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6): 559–572, 1901.

D. Peel and G. J. McLachlan. Robust mixture modelling using the t distribution. *Statistics and Computing*, 10(4):339–348, Oct. 2000.

H.-O. Peitgen, H. Jürgens, and D. Saupe. *Chaos and Fractals: New Frontiers of Science*. Springer-Verlag, New York, 1992.

J. Picone, S. Pike, R. Reagan, T. Kamm, J. Bridle, L. Deng, J. Ma, H. Richards, and M. Schuster. Initial evaluation of hidden dynamic models on conversational speech. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP'99)*, volume 1, pages 109–112, Phoenix, Arizona, USA, May 15–19 1999.

A. Pisani. A nonparametric and scale-independent method for cluster-analysis. 1. The univariate case. *Monthly Notices of the Royal Astronomical Society*, 265(3):706–726, Dec. 1993.

C. Posse. An effective two-dimensional projection pursuit algorithm. *Communications in Statistics — Simulation and Computation*, 19(4):1143–1164, 1990.

C. Posse. Tools for two-dimensional exploratory projection pursuit. *Journal of Computational and Graphical Statistics*, 4:83–100, 1995.

S. Pratt, A. T. Heintzelman, and D. S. Ensrud. The efficacy of using the IBM Speech Viewer vowel accuracy module to treat young children with hearing impairment. *Journal of Speech and Hearing Research*, 29: 99–105, 1993.

F. P. Preparata and M. I. Shamos. *Computational Geometry: An Introduction*. Monographs in Computer Science. Springer-Verlag, New York, 1985.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, U.K., second edition, 1992.

L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*. Signal Processing Series. Prentice-Hall, Englewood Cliffs, N.J., 1993.

M. G. Rahim, C. C. Goodyear, W. B. Kleijn, J. Schroeter, and M. M. Sondhi. On the use of neural networks in articulatory speech synthesis. *J. Acoustic Soc. Amer.*, 93(2):1109–1121, Feb. 1993.

R. A. Redner and H. F. Walker. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2):195–239, Apr. 1984.

K. Reinhard and M. Niranjan. Parametric subspace modeling of speech transitions. *Speech Communication*, 27(1):19–42, Feb. 1999.

M. Revow, C. K. I. Williams, and G. Hinton. Using generative models for handwritten digit recognition. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 18(6):592–606, June 1996.

H. B. Richards and J. S. Bridle. The HDM: a segmental hidden dynamic model of coarticulation. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP'99)*, volume I, pages 357–360, Phoenix, Arizona, USA, May 15–19 1999.

S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, B*, 59(4):731–758, 1997.

B. D. Ripley. *Pattern Recognition and Neural Networks.* Cambridge University Press, Cambridge, U.K., 1996.

S. J. Roberts. Parametric and non-parametric unsupervised cluster analysis. *Pattern Recognition*, 30(2): 261–272, Feb. 1997.

S. J. Roberts, D. Husmeier, I. Rezek, and W. Penny. Bayesian approaches to Gaussian mixture modeling. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 20(11):1133–1142, 1998.

W. J. J. Roberts and Y. Ephraim. Hidden Markov modeling of speech using Toeplitz covariance matrices. *Speech Communication*, 31(1):1–14, May 2000.

A. J. Robinson. An application of recurrent nets to phone probability estimation. *IEEE Trans. Neural Networks*, 5(2):298–305, Mar. 1994.

T. Rögnvaldsson. On Langevin updating in multilayer perceptrons. *Neural Computation*, 6(5):916–926, Sept. 1994.

R. Rohwer and J. C. van der Rest. Minimum description length, regularization, and multimodal data. *Neural Computation*, 8(3):595–609, Apr. 1996.

E. T. Rolls and A. Treves. *Neural Networks and Brain Function.* Oxford University Press, 1998.

K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proc. IEEE*, 86(11):2210–2239, Nov. 1998.

R. C. Rose, J. Schroeter, and M. M. Sondhi. The potential role of speech production models in automatic speech recognition. *J. Acoustic Soc. Amer.*, 99(3):1699–1709 (with comments, pp. 1710–1717), Mar. 1996.

E. Z. Rothkopf. A measure of stimulus similarity and errors in some paired-associate learning tasks. *J. of Experimental Psychology*, 53(2):94–101, 1957.

B. Rotman and G. T. Kneebone. *The Theory of Sets & Transfinite Numbers.* Oldbourne, London, 1966.

S. Roweis. EM algorithms for PCA and SPCA. In Jordan et al. (1998), pages 626–632.

S. Roweis. Constrained hidden Markov models. In Solla et al. (2000), pages 782–788.

S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 (5500):2323–2326, Dec. 22 2000.

A. E. Roy. *Orbital Motion.* Adam Hilger Ltd., Bristol, 1978.

D. B. Rubin. *Multiple Imputation for Nonresponse in Surveys.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1987.

D. B. Rubin and D. T. Thayer. EM algorithms for ML factor analysis. *Psychometrika*, 47(1):69–76, Mar. 1982.

D. B. Rubin and D. T. Thayer. More on EM for ML factor analysis. *Psychometrika*, 48(2):253–257, June 1983.

P. Rubin, T. Baer, and P. Mermelstein. An articulatory synthesizer for perceptual research. *J. Acoustic Soc. Amer.*, 70(2):321–328, Aug. 1981.

E. Saltzman and J. A. Kelso. Skilled actions: a task-dynamic approach. *Psychological Review*, 94(1):84–106, Jan. 1987.

J. W. Sammon, Jr. A nonlinear mapping for data structure analysis. *IEEE Trans. Computers*, C–18(5): 401–409, May 1969.

T. D. Sanger. Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473, 1989.

T. D. Sanger. Probability density estimation for the interpretation of neural population codes. *J. Neurophysiol.*, 76(4):2790–2793, Oct. 1996.

L. K. Saul and M. G. Rahim. Markov processes on curves. *Machine Learning*, 41(3):345–363, Dec. 2000a.

L. K. Saul and M. G. Rahim. Maximum likelihood and minimum classification error factor analysis for automatic speech recognition. *IEEE Trans. Speech and Audio Process.*, 8(2):115–125, Mar. 2000b.

E. Saund. Dimensionality-reduction using connectionist networks. *IEEE Trans. on Pattern Anal. and Machine Intel.*, 11(3):304–314, Mar. 1989.

J. A. Scales and M. L. Smith. *Introductory Geophysical Inverse Theory*. Samizdat Press, 1998. Freely available in draft form from `http://samizdat.mines.edu/inverse_theory/`.

F. Scarselli and A. C. Tsoi. Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results. *Neural Networks*, 11(1):15–37, Jan. 1998.

J. L. Schafer. *Analysis of Incomplete Multivariate Data*. Number 72 in Monographs on Statistics and Applied Probability. Chapman & Hall, London, New York, 1997.

B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors. *Advances in Kernel Methods. Support Vector Learning*. MIT Press, 1999a.

B. Schölkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Rätsch, and A. Smola. Input space vs. feature space in kernel-based methods. *IEEE Trans. Neural Networks*, 10(5):1000–1017, Sept. 1999b.

B. Schölkopf, A. Smola, and K.-R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, July 1998.

M. R. Schroeder. Determination of the geometry of the human vocal tract by acoustic measurements. *J. Acoustic Soc. Amer.*, 41(4):1002–1010, 1967.

J. Schroeter and M. M. Sondhi. Dynamic programming search of articulatory codebooks. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech and Sig. Proc. (ICASSP'89)*, volume 1, pages 588–591, Glasgow, UK, May 23–26 1989.

J. Schroeter and M. M. Sondhi. Techniques for estimating vocal-tract shapes from the speech signal. *IEEE Trans. Speech and Audio Process.*, 2(1):133–150, Jan. 1994.

M. Schuster. *On Supervised Learning from Sequential Data with Applications for Speech Recognition*. PhD thesis, Graduate School of Information Science, Nara Institute of Science and Technology, 1999.

D. W. Scott. *Multivariate Density Estimation. Theory, Practice, and Visualization*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1992.

D. W. Scott and J. R. Thompson. Probability density estimation in higher dimensions. In J. E. Gentle, editor, *Computer Science and Statistics: Proceedings of the Fifteenth Symposium on the Interface*, pages 173–179, Amsterdam, New York, Oxford, 1983. North Holland-Elsevier Science Publishers.

R. N. Shepard. Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, 5:33–48, 1963.

K. Shirai and T. Kobayashi. Estimating articulatory motion from speech wave. *Speech Communication*, 5(2): 159–170, June 1986.

M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch, editors. *Lévy Flights and Related Topics in Physics*. Number 450 in Lecture Notes in Physics. Springer-Verlag, Berlin, 1995. Proceedings of the International Workshop held at Nice, France, 27–30 June 1994.

B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Number 26 in Monographs on Statistics and Applied Probability. Chapman & Hall, London, New York, 1986.

L. Sirovich and M. Kirby. Low-dimensional procedure for the identification of human faces. *J. Opt. Soc. Amer. A*, 4(3):519–524, Mar. 1987.

D. S. Sivia. *Data Analysis. A Bayesian Tutorial*. Oxford University Press, New York, Oxford, 1996.

R. Snieder and J. Trampert. *Inverse Problems in Geophysics*. Samizdat Press, 1999. Freely available from `http://samizdat.mines.edu/snieder_trampert/`.

S. A. Solla, T. K. Leen, and K.-R. Müller, editors. *Advances in Neural Information Processing Systems*, volume 12, 2000. MIT Press, Cambridge, MA.

V. N. Sorokin. Determination of vocal-tract shape for vowels. *Speech Communication*, 11(1):71–85, Mar. 1992.

V. N. Sorokin, A. S. Leonov, and A. V. Trushkin. Estimation of stability and accuracy of inverse problem solution for the vocal tract. *Speech Communication*, 30(1):55–74, Jan. 2000.

C. Spearman. General intelligence, objectively determined and measured. *Am. J. Psychol.*, 15:201–293, 1904.

D. F. Specht. A general regression neural network. *IEEE Trans. Neural Networks*, 2(6):568–576, Nov. 1991.

M. Spivak. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. Addison-Wesley, Reading, MA, USA, 1965.

M. Spivak. *Calculus*. Addison-Wesley, Reading, MA, USA, 1967.

M. Stone. Toward a model of three-dimensional tongue movement. *J. of Phonetics*, 19:309–320, 1991.

N. V. Swindale. The development of topography in the visual cortex: A review of models. *Network: Computation in Neural Systems*, 7(2):161–247, May 1996.

A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 1987.

J. B. Tenenbaum. Mapping a manifold of perceptual observations. In Jordan et al. (1998), pages 682–688.

J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, Dec. 22 2000.

G. Tesauro, D. S. Touretzky, and T. K. Leen, editors. *Advances in Neural Information Processing Systems*, volume 7, 1995. MIT Press, Cambridge, MA.

R. J. Tibshirani. Principal curves revisited. *Statistics and Computing*, 2:183–190, 1992.

A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. Scripta Series in Mathematics. John Wiley & Sons, New York, London, Sydney, 1977. Translation editor: Fritz John.

M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analyzers. *Neural Computation*, 11(2):443–482, Feb. 1999a.

M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, B*, 61(3):611–622, 1999b.

D. M. Titterington, A. F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1985.

L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang. The indeterminacy and identifiability of blind identification. *IEEE Trans. Circuits and Systems*, 38(5):499–509, May 1991.

V. Tresp, R. Neuneier, and S. Ahmad. Efficient methods for dealing with missing data in supervised learning. In Tesauro et al. (1995), pages 689–696.

A. Treves, S. Panzeri, E. T. Rolls, M. Booth, and E. A. Wakeman. Firing rate distributions and efficiency of information transmission of inferior temporal cortex neurons to natural visual stimuli. *Neural Computation*, 11(3):601–632, Mar. 1999.

A. Treves and E. T. Rolls. What determines the capacity of autoassociative memories in the brain? *Network: Computation in Neural Systems*, 2(4):371–397, Nov. 1991.

A. C. Tsoi. Recurrent neural network architectures — an overview. In C. L. Giles and M. Gori, editors, *Adaptive Processing of Temporal Information*, volume 1387 of *Lecture Notes in Artificial Intelligence*, pages 1–26. Springer-Verlag, New York, 1998.

UCLA. Artificial EPG palate image. The UCLA Phonetics Lab. Available online at `http://www.humnet.ucla.edu/humnet/linguistics/faciliti/facilities/physiology/EGP_picture.JPG`, Feb. 1, 2000.

N. Ueda, R. Nakano, Z. Ghahramani, and G. E. Hinton. SMEM algorithm for mixture models. *Neural Computation*, 12(9):2109–2128, Sept. 2000.

A. Utsugi. Hyperparameter selection for self-organizing maps. *Neural Computation*, 9(3):623–635, Apr. 1997a.

A. Utsugi. Topology selection for self-organizing maps. *Network: Computation in Neural Systems*, 7(4): 727–740, 1997b.

A. Utsugi. Bayesian sampling and ensemble learning in generative topographic mapping. *Neural Processing Letters*, 12(3):277–290, Dec. 2000.

A. Utsugi and T. Kumagai. Bayesian analysis of mixtures of factor analyzers. *Neural Computation*, 13(5): 993–1002, May 2001.

V. N. Vapnik and S. Mukherjee. Support vector method for multivariate density estimation. In Solla et al. (2000), pages 659–665.

S. V. Vaseghi. *Advanced Signal Processing and Digital Noise Reduction*. John Wiley & Sons, New York, London, Sydney, second edition, 2000.

S. V. Vaseghi and P. J. W. Rayner. Detection and suppression of impulsive noise in speech-communication systems. *IEE Proc. I (Communications, Speech and Vision)*, 137(1):38–46, Feb. 1990.

T. Villmann, R. Der, M. Hermann, and T. M. Martinetz. Topology preservation in self-organizing feature maps: Exact definition and measurement. *IEEE Trans. Neural Networks*, 8(2):256–266, Mar. 1997.

W. E. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–1276, Feb. 18 2000.

H. M. Wagner. *Principles of Operations Research with Applications to Managerial Decisions*. Prentice-Hall, Englewood Cliffs, N.J., second edition, 1975.

A. Webb. *Statistical Pattern Recognition*. Edward Arnold, 1999.

A. R. Webb. Multidimensional scaling by iterative majorization using radial basis functions. *Pattern Recognition*, 28(5):753–759, May 1995.

E. J. Wegman. Hyperdimensional data analysis using parallel coordinates. *J. Amer. Stat. Assoc.*, 85(411): 664–675, Sept. 1990.

J. R. Westbury. *X-Ray Microbeam Speech Production Database User's Handbook Version 1.0*. Waisman Center on Mental Retardation & Human Development, University of Wisconsin, Madison, WI, June 1994. With the assistance of Greg Turner & Jim Dembowski.

J. R. Westbury, M. Hashi, and M. J. Lindstrom. Differences among speakers in lingual articulation for American English /ɹ/. *Speech Communication*, 26(3):203–226, Nov. 1998.

J. Weston, A. Gammerman, M. O. Stitson, V. Vapnik, V. Vovk, and C. Watkins. Support vector density estimation. In Schölkopf et al. (1999a), chapter 18, pages 293–306.

J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York, London, Sydney, 1990.

P. Whittle. On principal components and least square methods of factor analysis. *Skand. Aktur. Tidskr.*, 36: 223–239, 1952.

J. Wiles, P. Bakker, A. Lynton, M. Norris, S. Parkinson, M. Staples, and A. Whiteside. Using bottlenecks in feedforward networks as a dimension reduction technique: An application to optimization tasks. *Neural Computation*, 8(6):1179–1183, Aug. 1996.

J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, New York, Oxford, 1965.

P. M. Williams. Using neural networks to model conditional multivariate densities. *Neural Computation*, 8 (4):843–854, May 1996.

B. Willmore, P. A. Watters, and D. J. Tolhurst. A comparison of natural-image-based models of simple-cell coding. *Perception*, 29(9):1017–1040, Sept. 2000.

R. Wilson and M. Spann. A new approach to clustering. *Pattern Recognition*, 23(12):1413–1425, 1990.

J. H. Wolfe. Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5:329–350, July 1970.

D. M. Wolpert and Z. Ghahramani. Computational principles of movement neuroscience. *Nat. Neurosci.*, 3 (Supp.):1212–1217, Nov. 2000.

D. M. Wolpert and M. Kawato. Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7–8):1317–1329, Oct. 1998.

A. A. Wrench. A multi-channel/multi-speaker articulatory database for continuous speech recognition research. In *Phonus*, volume 5, Saarbrücken, 2000. Institute of Phonetics, University of Saarland.

F. Xie and D. van Compernolle. Speech enhancement by spectral magnitude estimation —a unifying approach. *Speech Communication*, 19(2):89–104, Aug. 1996.

L. Xu, C. C. Cheung, and S. Amari. Learned parametric mixture based ICA algorithm. *Neurocomputing*, 22 (1–3):69–80, Nov. 1998.

E. Yamamoto, S. Nakamura, and K. Shikano. Lip movement synthesis from speech based on hidden Markov models. *Speech Communication*, 26(1–2):105–115, 1998.

H. H. Yang and S. Amari. Adaptive on-line learning algorithms for blind separation — maximum entropy and minimum mutual information. *Neural Computation*, 9(7):1457–1482, Oct. 1997.

H. Yehia and F. Itakura. A method to combine acoustic and morphological constraints in the speech production inverse problem. *Speech Communication*, 18(2):151–174, Apr. 1996.

H. Yehia, T. Kuratate, and E. Vatikiotis-Bateson. Using speech acoustics to drive facial motion. In Ohala et al. (1999), pages 631–634.

H. Yehia, P. Rubin, and E. Vatikiotis-Bateson. Quantitative association of vocal-tract and facial behavior. *Speech Communication*, 26(1–2):23–43, Oct. 1998.

G. Young. Maximum likelihood estimation and factor analysis. *Psychometrika*, 6:49–53, 1940.

S. J. Young. A review of large vocabulary continuous speech recognition. *IEEE Signal Processing Magazine*, 13(5):45–57, Sept. 1996.

K. Zhang, I. Ginzburg, B. L. McNaughton, and T. J. Sejnowski. Interpreting neuronal population activity by reconstruction: Unified framework with application to hippocampal place cells. *J. Neurophysiol.*, 79(2): 1017–1044, Feb. 1998.

R. D. Zhang and J.-G. Postaire. Convexity dependent morphological transformations for mode detection in cluster-analysis. *Pattern Recognition*, 27(1):135–148, 1994.

Y. Zhao and C. G. Atkeson. Implementing projection pursuit learning. *IEEE Trans. Neural Networks*, 7(2): 362–373, Mar. 1996.

I. Zlokarnik. Adding articulatory features to acoustic features for automatic speech recognition. *J. Acoustic Soc. Amer.*, 97(5):3246, May 1995a.

I. Zlokarnik. Articulatory kinematics from the standpoint of automatic speech recognition. *J. Acoustic Soc. Amer.*, 98(5):2930–2931, Nov. 1995b.