# Multivariate Cepstral Feature Compensation on Band-limited Data for Robust Speech Recognition

**Nicolas Morales**
HCTLab
Univ. Autónoma Madrid
Madrid. Spain
nicolas.morales@uam.es

**Doroteo T. Toledano**
ATVSLab
Univ. Autónoma Madrid
Madrid. Spain
doroteo.torre@uam.es

**John H. L. Hansen**
CRSS
Univ. of Texas at Dallas
Richardson, TX. USA
john.hansen@utdallas.edu

**Javier Garrido**
HCTLab
Univ. Autónoma Madrid
Madrid. Spain
javier.garrido@uam.es

## Abstract

This paper describes a new method for compensating bandwidth mismatch for automatic speech recognition using multivariate linear combinations of feature vector components. It is shown that multivariate compensation is superior to methods based on linear compensations of individual features. Performance is evaluated on a real microphone-telephone mismatch condition (this involves noise compensation and bandwidth extension of real data), as well as on several artificial bandwidth limitations. Speech recognition accuracy using this approach is similar to that of acoustic model compensation methods for small to moderate mismatches, and allows keeping active a single acoustic model set for multiple bandwidth limitations.

## 1 Introduction

Noise robustness is a major issue in current research on Automatic Speech Recognition (ASR). Systems trained and tested under laboratory conditions reach high accuracy rates. However, when there is a mismatch between training and test conditions accuracy is severely affected.

This work studies the problem of mismatch between training and test in terms of available frequency bandwidth. Speech recognition systems are typically trained on full-bandwidth data (for speech recognition systems this is normally 0-8kHz). However, in real implementations part of the spectrum of input data could be missing; for example, this situation could be created by a channel distortion or sampling frequency below 16kHz.

Clearly, a simple solution to this problem is retraining new models for the specific type of channel. However, it may well be the case that not enough training data is available from the new environment. Also, when a wide range of possible band-limitations exists for a particular application training of acoustic models for each of them is not appropriate.

Our approach is to compensate band-limited feature vectors to generate pseudo-full-bandwidth features that can be passed to a speech recognizer trained on full-bandwidth speech. The advantages are twofold: first, it is easy to train and requires only small amounts of data. Second, the recognizer module keeps a single acoustic recognizer active at all times, a desirable situation for small devices where memory limitation and energy consumption are relevant.

Feature compensation has been used in the past, especially for speech affected by noise (Moreno, 1996; Droppo et al., 2001). In other cases, compensation is introduced in the decoder module (Deng et al., 2005).

For the case of bandwidth mismatch feature compensation has recently been used in the form of univariate linear and polynomial correction (Seltzer et al., 2005; Morales et al., 2005). These studies proposed compensation directly in the domain of Mel Frequency Cepstrum Coefficients

**Joakim Nivre, Heiki-Jaan Kaalep, Kadri Muischnek and Mare Koit (Eds.)**
**NODALIDA 2007 Conference Proceedings, pp. 144–151**

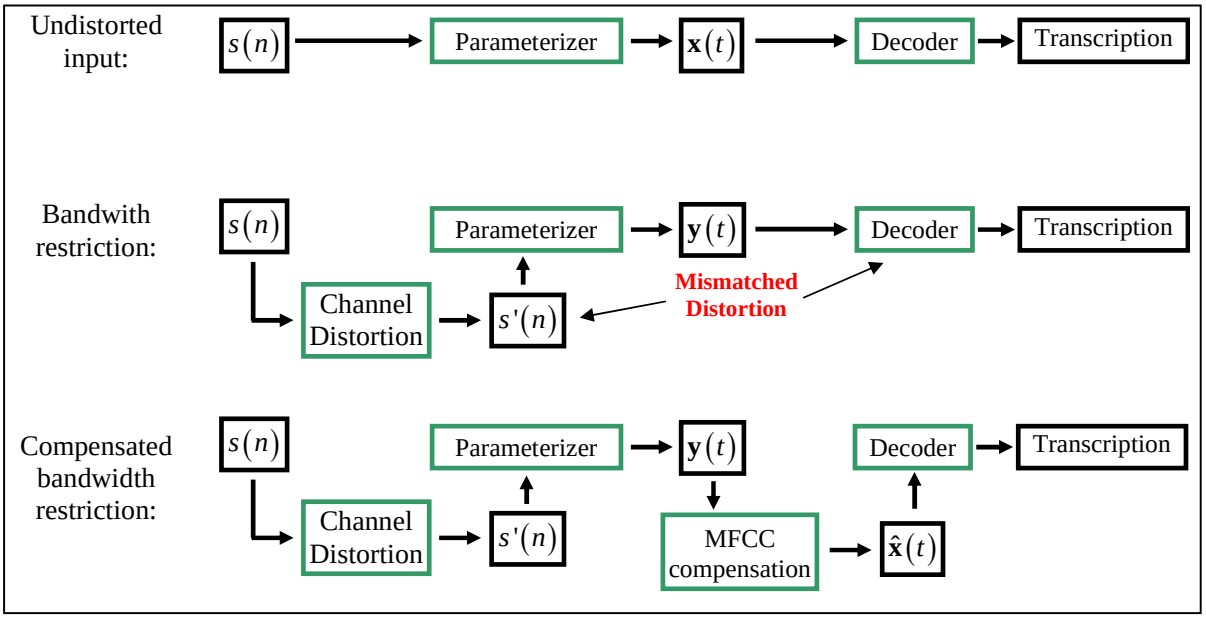


**Figure 1.** Modification of a basic speech recognizer system for MFCC feature compensation. The ideal working environment is noted as "Undistorted input". However, in many cases, some kind of distortion, affects the input signal, producing a mismatch between the characteristics of speech and the acoustic models of the decoder (in our case a bandwidth restriction). In this study mismatch is reduced by introducing an MFCC compensation module between the parameterizer and decoder modules.

(MFCC), the parameterization of choice for most speech recognizers. Thus, the compensator module may be easily inserted between the parameterizer and recognizer modules of already working ASR engines (Figure 1).

In this work we propose the use of multivariate linear correction for bandwidth compensation. Each individual MFCC is compensated using a linear combination of a selection of other coefficients in the same frame. The previously referenced univariate compensation algorithms corrected each MFCC coefficient independently based on the assumption that MFCCs are highly uncorrelated. However, as we show in Section 3, this assumption is less valid when data is band-limited. Experiments show that by discarding this assumption, better compensation and ASR performance may be achieved.

Band-restricted speech can be found in historical spoken document retrieval (Hansen et al., 2004). In broadcast news' transcription it may also occur that the channel conditions change abruptly and rapidly, for example when the studio presenter talks to an anchor in a foreign country. Other cases where multiple band-limiting distortions may be found are on-board systems, such as those in cars, or in airplanes (Abut et al., 2005; Denenberg et al., 1993). In these cases using multiple acoustic model sets for the different conditions could be costly and complicated. On the contrary, feature compensation generalizes seamlessly to such aggressive environments; for example, it has been shown that multiple band-limitations may be automatically classified and successfully compensated using a single compensation system, and also that data from a sufficient number of environments allows for compensation of unseen distortions (Morales et al., 2007). These properties are related to the method employed for partitioning the limited-bandwidth MFCC space (Section 4) and are independent of whether univariate or multivariate compensations are applied. Thus, they hold true for multivariate compensation.

The rest of the paper is organized as follows: Section 2 introduces MFCC compensation and Section 3 discusses on the need of multivariate compensation for band-limited speech. Section 4 describes practical issues and Section 5 presents experimental results. In Section 6 conclusions are presented.

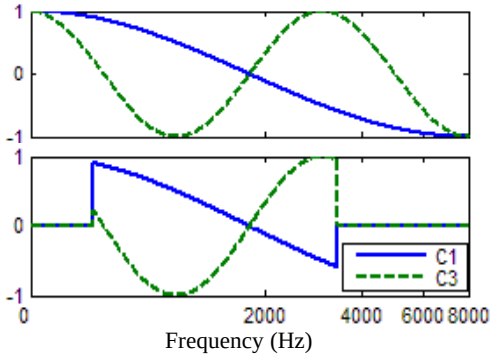


**Figure 2.** Cepstral transforms of orders 1 and 3 for full-bandwidth (top) and limited-bandwidth speech (bottom; 300-3400Hz band-pass filter). Band-limited transforms are no longer orthogonal.

## 2 MFCC Compensation

Previous works have studied in detail the effect of band-limiting distortions on the MFCCs (Huang et al., 2001; Morales et al., 2005). Here, we present their main conclusions.

The band-limited MFCC space may be modeled as a mixture of $K$ Gaussian classes:

$$p(\mathbf{y}) = \sum_{k=1}^{K} N\left(\mathbf{y}; \boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k\right) \cdot P(k), \qquad (1)$$

where $\mathbf{y}$ is the band-limited feature vector and $N\left(\square; \boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k\right)$ is the Gaussian distribution with mean vector $\boldsymbol{\mu}^k$ and covariance matrix $\boldsymbol{\Sigma}^k$ associated to class $k$. The full-bandwidth space is modeled similarly and assuming that both spaces are jointly Gaussian for each class $k$, the expectation of the full-bandwidth vector $\mathbf{x}$ is:

$$\hat{\mathbf{x}}(\mathbf{y}, k) = E\{\mathbf{x}|\mathbf{y}, k\} = \boldsymbol{\mu}_{\mathbf{x}}^k + \boldsymbol{\Sigma}_{\mathbf{xy}}^k \left(\boldsymbol{\Sigma}_{\mathbf{y}}^k\right)^{-1} \left(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}}^k\right) = \\ = \mathbf{B}^k \mathbf{y} + \mathbf{b}^k, \qquad (2)$$

where $\mathbf{B}^k$ and $\mathbf{b}^k$ are the compensation matrix and offset vector for class $k$, and sub-indexes $\mathbf{x}$ and $\mathbf{y}$ indicate full-bandwidth or limited bandwidth speech, respectively. Generally, the importance of non-diagonal terms was assumed negligible and $\mathbf{B}^k$ was diagonalized (Droppo et al., 2001; Morales et al., 2005). Thus, an expression for individual full-bandwidth MFCC coefficients may be simplified from (2) as:

$$x_i \approx \hat{x}_i\left(y_i | k\right) = B_i^k \cdot y_i + b_i^k, \qquad (3)$$

where $i$ is the order of the MFCC coefficient, $b_i^k$ is element $i$ of vector $\mathbf{b}^k$ and $B_i^k$ the diagonal element $(i, i)$ in matrix $\mathbf{B}^k$.

As will be shown in the following section, the diagonal simplification in (3) that is acceptable on full-bandwidth speech corrupted by noise could be harmful when it is applied to band-limited speech.

## 3 On MFCC Uncorrelation and Band-limiting Distortions

MFCC features are generally assumed uncorrelated. In fact, this is one of the key points for their extended use in ASR systems – they allow using diagonal covariance matrices in Gaussian mixture models without significant performance loss. In the past, this assumption led to the use of diagonal compensation matrices for MFCC feature compensation. However, we recently observed that MFCC features coming from band-limited speech showed a higher degree of correlation than those coming from full-bandwidth speech.

In order to compare the degree of correlation between MFCC parameters we defined the following measure of non-diagonality for the covariance matrix:

$$nonDiag = \sum_{i}^{staticMFCCs} \sum_{j,\, j \neq i}^{MFCCs} \delta_{ij},$$

$$\delta_{ij} = \begin{cases} 1 & if\ \sqrt{\mathrm{cov}(i,i) \cdot \mathrm{cov}(j,j)} \leq \tau \cdot \mathrm{cov}(i,j) \\ 0 & otherwise \end{cases}. \qquad (4)$$

Using $\tau = 5$ we obtained a nonDiagonality of 51 for full-bandwidth features, 108 for their corresponding 4kHz low-pass filtered features and 110 for a band-pass filter 300-3400Hz (similar results are found with other values of $\tau$). This shows that filtered MFCCs are more correlated than full-bandwidth MFCCs. Thus, the general assumption of uncorrelation seems less valid for band-limited MFCCs and the use of a non-diagonal compensation matrix is justified.

From (2) we can establish the relationship between the covariance matrices of band-limited and full-bandwidth MFCCs as:

$$\boldsymbol{\Sigma}_{\mathbf{x}}^k = \mathbf{B}^k \cdot \boldsymbol{\Sigma}_{\mathbf{y}}^k \cdot \left(\mathbf{B}^k\right)^t. \qquad (5)$$

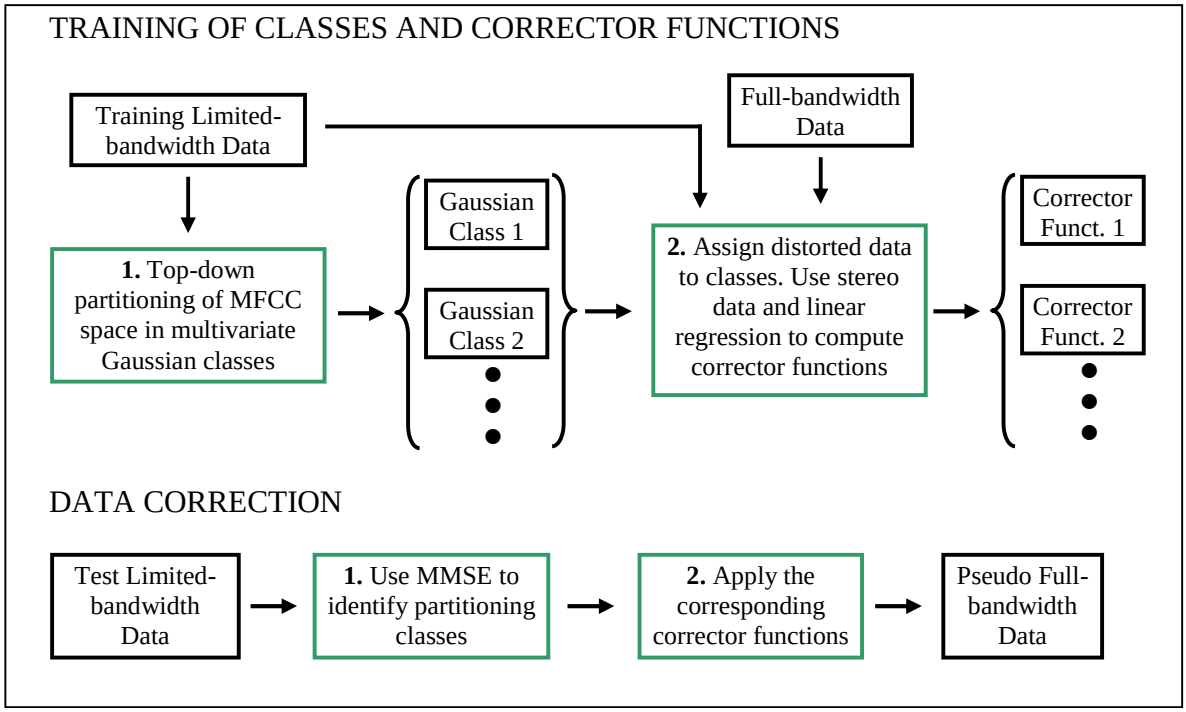


**Figure 3. S**chematic representations of the proposed architectures for training of classes and corrector functions and for compensation of band-limited MFCCs to generate pseudo-full bandwidth MFCCs.

Assuming that the covariance matrix of full-bandwidth MFCCs, $\mathbf{\Sigma}_{\mathbf{x}}^{k}$, is diagonal and that of band-limited features, $\mathbf{\Sigma}_{\mathbf{y}}^{k}$, is non-diagonal, then the compensation matrix, $\mathbf{B}^{k}$ needs to be non-diagonal, in order to satisfy (5).

The approximately uncorrelated nature of MFCCs has been empirically observed on speech data and is associated with the fact that the Discrete Cosine Transform (DCT) on filterbank energies and Principal Component Analysis (PCA) on the correlation matrix generate very similar transformations (Pols, 1977). However, as seen in Figure 2, using the DCT on band-limited frames is effectively a different transformation of that over full-bandwidth speech. The vectors in the basis are no longer orthogonal (on the contrary DCT on full-bandwidth data as well as PCA are orthogonal transforms) and empirical evidence suggests that this could increase correlation of band-limited MFCCs compared to full-bandwidth features (though more experiments should be done for better comprehension of this phenomenon).

Because our compensation framework does not require matrix inversions or expensive calculations the computational cost of non-diagonal compensation matrices may be assumed if, as will be shown later, significant performance gains may be achieved.

## 4 Class and Corrector Function Training

The proposed framework is shown in Figure 3. Training consists of two steps. First, the partitioning classes from each environment are created and second, a corrector function is computed for each class and MFCC feature. When a system needs to be deployed in an environment where different types of bandwidth limitations may exist, classes and corrector functions are created independently for each of the existing conditions. Classes trained with data from the different distortions will be able to identify the type of distortion of incoming data and will then apply the appropriate compensation functions. Also, if the need to create classes for new distortions arises, these can be added to the existing framework without any further modification (Morales et al., 2007).

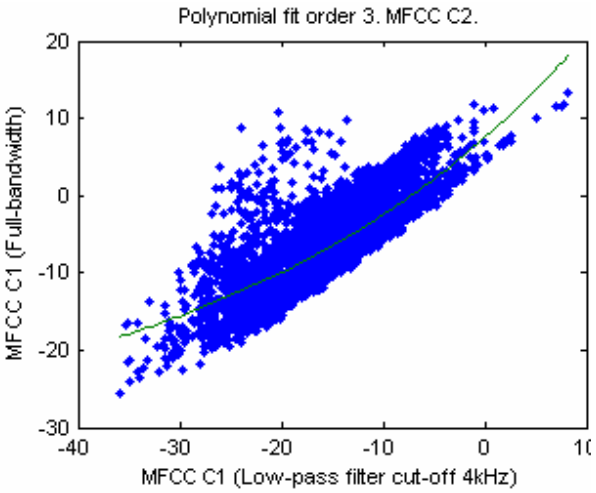


**Figure 4.** Mapping of low-pass filter 4kHz data to full-bandwidth for MFCC parameter C2 in a particular class $k$. The plot also shows a third order polynomial fit.

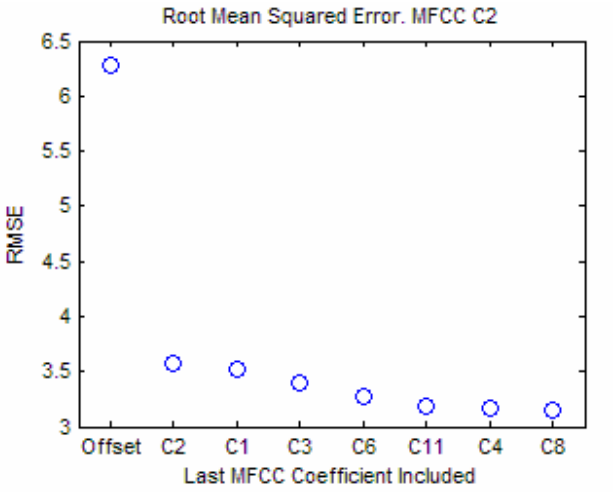


**Figure 5.** Root Mean Squared Error (RMSE) for multivariate fit of full-bandwidth MFCC C2 in a particular class $k$ of the limited bandwidth space (for a low-pass filter, cut-off frequency 4kHz). RMSE improves as more coefficients are included in the fit. Ticks in the x-axis show the best coefficient to add in each step (C2, C1, etc. indicate static MFCC coefficients of orders 2, 1, etc., respectively).

### 4.1 Class Creation

For each target distorting environment a different set of Gaussian classes is generated using a top-down approach: an initial multivariate Gaussian distribution with mean and diagonal covariance computed from all the training data is divided into two classes. Data are then re-assigned to either class and their mean vector and covariance matrix are re-estimated. The process is repeated introducing new classes in successive iterations until the number of final mixtures is reached.

### 4.2 Corrector Function Training

Separate correction matrices and offset vectors are trained for each compensation class defined in the restricted-bandwidth space as explained in Section 4.1. In our experiments we use stereo data to compute the coefficients in the corrector functions (here stereo data refers to speech recorded simultaneously under the full-bandwidth and limited bandwidth environments. Alternatively, when a good characterization of the distortion is available it is possible to generate pseudo-distorted data).

Band-limited speech frames from the training set are assigned to one of the corrector classes previously defined based on a maximum likelihood criterion:

$$\hat{k}(t) = \max_k \left( N\left(\mathbf{y}_t; \boldsymbol{\mu}^k, \boldsymbol{\Sigma}^k\right) \cdot P(k) \right), \quad 1 \le k \le K, \quad (6)$$

where $K$ is the total number of classes. For univariate polynomial correction, each MFCC value in the band-limited space is mapped to its equivalent in the full-bandwidth space. In Figure 4 each point represents the value of a given MFCC in the band-limited space (x-axis) and full-bandwidth space (y-axis). Then, for each corrector class the corresponding $\mathbf{B}^k$ and $\mathbf{b}^k$ are computed using linear regression (the green curve in Figure 4). For multivariate linear correction a similar approach is followed identifying feature vectors from stereo frames in the full-bandwidth and limited-bandwidth spaces and employing multivariate linear regression. Multivariate linear regression follows an incremental form, starting from a simple offset and adding successively the coefficient for which a higher decrease of Mean Squared Error (MSE) is achieved until no significant decrease is found. In this way, it is possible to determine the ideal number of MFCC coefficients to use for the compensation of a particular component. In figure 5 we show explicitly the evolution of the Root Mean Squared Error (RMSE) after inclusion of each individual coefficient in the regression. The target coefficient is full-bandwidth MFCC C2 and not surprisingly the first coefficient inserted is limited-bandwidth MFCC C2. Going from a simple

offset to compensation with a single coefficient reduces RMSE from 6.28 to 3.58. This is equivalent to univariate linear compensation. However, the inclusion of the next 6 coefficients (C1, C3, C6, C11 and C4) further reduces RMSE to 3.14, which seems to indicate that significant benefits may be obtained by applying multivariate compensation. On the contrary inclusion of additional coefficients offers very little improvement, which indicates that in this case, compensation may be truncated after the best 7 coefficients.

Data compensation uses an MMSE version of (2) for multivariate and (3) for univariate compensation.

## 5 Results and Discussion

Experiments are based on two measures: first, direct reconstruction quality is assessed by computing the average Mahalanobis distance between real full-bandwidth data and estimated pseudo-full-bandwidth data (generated by compensation of limited-bandwidth data); second, ASR accuracy is evaluated using full-bandwidth acoustic models on pseudo-full-bandwidth data.

### 5.1 Measuring Reconstruction Quality

The quality of feature compensation may be directly measured in terms of a distance metric between the real full-bandwidth vectors and their corresponding reconstructed vectors. The ultimate goal being ASR performance, perfect reconstruction of feature vectors may be unnecessary as long as speech recognition decoding performs satisfactorily. However, a direct measure is useful because it is fast and independent of external elements such as grammar, phoneme list or other tunable parameters.

The quality measure used in this work is the average Mahalanobis distance. Table 1 shows a comparison between univariate linear compensation (*Univar*) and multivariate linear compensation (*Multivar*). As can be seen, multivariate linear compensation offers better performance for each group of MFCC parameters (this holds for each individual parameter, though a full table is not presented here for lack of space). We also compare reconstruction of dynamic parameters using feature compensation (*Multivar dynamic*) or computation with the typical definition of dynamic features, i.e. using linear regression on reconstructed static features (*Multivar static*). Not surprisingly, the distance is smaller using *Multivar dynamic* compensation, because feature compensation minimizes MSE between the actual full-bandwidth data and pseudo-full-bandwidth data. However, from the point of view of speech recognition accuracy we have observed that dynamic features computed by regression of static features (*Multivar static*) is better. Thus, it seems that even if the actual MSE is minimized using feature compensation for dynamic features, this may cause incongruence between static and dynamic features producing a loss in accuracy (for example, in the case of low-pass filter with cut-off frequency 4kHz, regression obtains a relative 0.76% accuracy gain compared to dynamic feature compensation).

| Mahalanobis Dist. ($x10^{-2}$) | Univar static | Multivar dynamic | Multivar static |
|---|---|---|---|
| **Static MFCCs** | 0.7848 | 0.7091 | 0.7091 |
| **Δ MFCCs** | 0.8180 | 0.7193 | 0.7234 |
| **ΔΔ MFCCs** | 0.8582 | 0.7393 | 0.7526 |
| **Total** | **2.461** | **2.168** | **2.185** |
| **ASR accuracy** | 66.97 | 68.22 | 68.46 |

**Table 1.** Mahalanobis distance between real full-bandwidth data and reconstructed data from low-pass filtered data with cut-off frequency 4kHz.

### 5.2 Measuring Speech Recognition

Speech recognition of reconstructed speech is evaluated using a phonetic recognition engine based on 51 Hidden Markov Models (HMM) and a phone bigram. The front-end uses pre-emphasis filtering ($\alpha$=0.97) and 25ms Hamming windows with a 10ms window shift. Thirteen MFCC coefficients including C0 and their respective first and second order derivatives (39 total features) are computed from a filter-bank of 26 Mel-scaled filters distributed in the region 0-8 kHz. HMM models are trained using TIMIT (Fisher et al., 1986). For training we use all 4680 files in the training partition and evaluation is made on all the 1620 files in the test partition.

#### Comparison of Different Approaches

In this section different approaches are considered for the problem of band-limited input speech. Table 2 shows results for artificial filters applied on TIMIT: Low-Pass 6kHz, Low-Pass 4kHz and Band-Pass 300-3400Hz, the last one simulating a

| Test Set | Correction | Percent Correct | Percent Accuracy |
|---|---|---|---|
| Full-Band | None | 75.40 | 71.18 |
| Low- Pass 6kHz | None | 64.32 | 58.30 |
| | Matched | 75.45 | 71.03 |
| | Model Adapt | 74.97 | 70.35 |
| | Univariate-32 | 74.88 | 70.65 |
| | Multivariate-32 | 75.22 | 70.95 |
| Low- Pass 4kHz | None | 55.93 | 44.67 |
| | Matched | 74.73 | 69.33 |
| | Model Adapt | 73.30 | 68.38 |
| | Univariate-32 | 72.41 | 66.97 |
| | Multivariate-32 | 73.16 | 68.46 |
| Band- Pass 300-3400 Hz | None | 41.13 | 32.67 |
| | Matched | 71.86 | 65.73 |
| | Model Adapt | 70.04 | 64.25 |
| | Univariate-32 | 65.63 | 58.46 |
| | Multivariate-32 | 69.29 | 63.44 |
| Real telephone data | None | 30.98 | 21.23 |
| | Matched | 69.10 | 61.80 |
| | Model Adapt | 66.86 | 61.22 |
| | Univariate-32 | 56.03 | 49.14 |
| | Univariate-256 | 60.32 | 53.38 |
| | Multivariate-32 | 62.53 | 56.78 |
| | Multivariate-256 | 64.67 | 58.79 |

**Table 2.** Band-limited speech recognition results. In Univariate and Multivariate the number that follows indicates the amount of classes employed for band-limited space partitioning.

noise-free telephone channel. In addition, performance on real telephone data is given: the whole TIMIT database was passed through the telephone line in a single call. This is similar to NTIMIT (Jankowski et al., 1990), but in our case all data is distorted by the same channel; a desirable condition in stereo-data compensation.

For comparison, results are given in the first row for the case of full-bandwidth training and test data, setting the upper limit performance. Recognition with full-bandwidth models and restricted-bandwidth test data incurs in a significant accuracy loss even for small distortions like a 6kHz low-pass filter (accuracy goes from 71.18% to 58.30%, a relative 45% error increase; see Table 2). Thus, some compensation (either on the feature or the model side) needs to be applied.

The new multivariate linear correction approach clearly and significantly outperforms polynomial correction showing the convenience of a non-diagonal matrix for feature compensation (i.e. multivariate compensation). Also, the performance achieved is similar to that of model compensation approaches, even for the real telephone distortion,

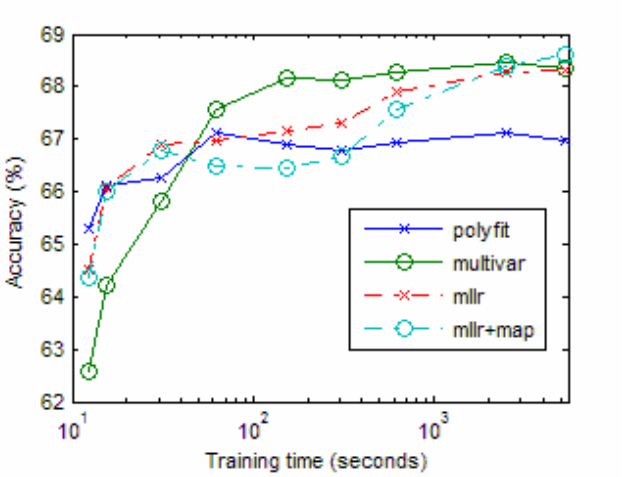


**Figure 6.** Accuracy for different feature compensation and model-based approaches for 8kHz-4kHz mismatch vs. available training data (in seconds).

were multivariate compensation is only 2.4% absolute worse than with model adaptation.

An important consideration is the number of corrector classes to be used. Previous experiments showed how compensation performance saturates for a *large* number of classes. Dealing with artificial filters, saturation appears for a number of classes around 25 (in our experiments, only 32 classes were used). On the contrary, for the more complicated situation of real telephone data, where noise is also present, a larger number of classes produced a very substantial improvement (compare results for 32 and 256 classes for this case).

### Limited Amounts of Training Data

In real applications it could be difficult to produce sufficient amounts of training material for feature compensation or model adaptation. Figure 6 shows performance relative to the amount of training data available. *MLLR* denotes global MLLR adaptation followed by 32-class MLLR adaptation. *MLLR+MAP* uses MAP adaptation on previously MLLR-adapted models (this is also used for model adaptation in Table 2). When the amount of training material is very limited, model adaptation outperforms multivariate compensation, showing the effectiveness of global MLLR (the first stage applied in model adaptation). However, the learning slope in multivariate feature compensation is steeper and from ~50 seconds of training material, multivariate linear correction obtains better results

than model adaptation methods, remaining so for as much as ~40 minutes of speech. Thus, it seems that at least for this particular case of filtering distortions and limited data, feature compensation could be a better approach than model adaptation.

## 6 Conclusions

A new feature compensation framework based on multivariate linear correction was presented. Feature compensation for robust ASR under multiple distorting environments is desirable because it allows using a single acoustic model set independent of the number of distorting environments, and keeps memory load and computation requirements low.

ASR accuracy with the proposed algorithm is similar to that of model-compensation approaches if large amounts of training material are available. In addition, when the amount of training data is small, multivariate linear correction shows better accuracy than all the other approaches considered. Experiments on real telephone data where also conducted showing very promising results (only ~2% absolute loss compared to model adaptation).

The new approach clearly outperforms our previous polynomial compensation with very small increase in computation time. This shows the great advantage of a full compensation matrix over a diagonal one for the case of band-limited data and is in agreement with the practical observations in Sections 3 and 4.2.

In the future, the need of stereo data should be overcome to allow straightforward application to a variety of new practical situations.

## Acknowledgments

This research is supported in part by an MCyT project (TIC 2006-13141-C03).

## References

H. Abut, J.H.L Hansen and K. Takeda (eds.). 2005. *DSP for in-vehicle and mobile systems*. Kluwer/Springer-Verlag.

L. Denenberg, H. Gish, M. Meteer, T. Miller, J.R. Rohlicek, W. Sadkin and M. Siu. 1993. Gisting conversational speech in real time. *Proceedings ICASSP, 2: 131-134*.

L. Deng, J. Droppo and A. Acero. 2005. Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion. *IEEE Speech and Audio Processing, 13(3):412-421*.

J. Droppo, L. Deng and A. Acero. 2001. Evaluation of the SPLICE algorithm on the Aurora2 database. *Proceedings EuroSpeech, 217-220*.

W. M. Fisher, R. Doddington and K. M. Goudie-Marshall. 1986. The DARPA Speech Recognition Research Database: Specifications and Status. *Proceedings DARPA Workshop on Speech Recognition, 93-99*.

J. H. L. Hansen, R. Huang, P. Mangalath, B. Zhou, M. Seadle, M. and J. Deller. 2004. SPEECHFIND: spoken document retrieval for a national gallery of the spoken word. *NORSIG, 1-4*.

X. Huang, A. Acero and H. W. Hon. 2001. *Spoken language processing*. Prentice Hall.

C. Jankowski, A. Kalyanswamy, S. Basson and J. Spitz. 1990. NTIMIT: A Phonetically Balanced, Continuous Speech, Telephone Bandwidth Speech Database. *Proceedings of ICASSP, 1:109-112* .

N. Morales, D. T. Toledano, J. H. L. Hansen, J. Colas and J. Garrido. 2005. Statistical class-based MFCC enhancement of filtered and band-limited speech for robust ASR. *Proceedings EuroSpeech, 2629-2632*.

N. Morales, D. T. Toledano, J. H. L. Hansen and J. Colas. 2007. Blind feature compensation for time-variant band-limited speech recognition. *IEEE Signal Processing Letters*, 14(1):70-73.

P. Moreno. 1996. *Speech recognition in noisy environments*. PhD. Thesis in Electrical and Computer Engineering. Carnegie Mellon University, Pittsburgh.

L. C. W. Pols. 1977. *Spectral analysis and identification of Dutch vowels in monosyllabic words*. Ph.D. Thesis. Free University of Amsterdam.

M. Seltzer, A. Acero and J. Droppo. 2005. Robust bandwidth extension of noise-corrupted narrowband speech. *Proceedings EuroSpeech, 1509-1512*.