Study of Probabilistic and Bottle-Neck Features in Multilingual Environment

František Grézl, Martin Karafiát, Miloš Janda

Brno University of Technology, Speech@FIT
Božetěchova 2, 612 66 Brno, Czech Republic
{grezl,karafiat,ijanda}@fit.vutbr.cz

***Abstract*—This study is focused on the performance of Probabilistic and Bottle-Neck features on different language than they were trained for. It is shown, that such porting is possible and that the features are still competitive to PLP features. Further, several combination techniques are evaluated. The performance of combined features is close to the best performing system. Finally, bigger NNs were trained on large data from different domain. The resulting features outperformed previously trained systems and combination with them further improved the system performance.**

I. Introduction

The increasing interest in speech-to-speech translation and automatic processing of low-resource languages led to research of multilingual approaches which would ease the system development for a new language. The biggest cost factor in such development is the need of training data for the acoustic model. Several techniques have been investigated to alleviate this problem. The *cross-language transfer* applies a system developed on one language to another one. It has been shown that the performance in new language is proportional to the similarity of the languages [1]. The *language adaptation* technique adapts the system to a new language with only limited data. The performance of the adapted system depends on the amount of available data [2]. When the amount of data becomes sufficient for full training, the *bootstrapping* technique can be used for initializing the new language system by the original one.

But having low-cost monolingual systems might not solve the problem completely. To process a recording with unknown language, it would be necessary to perform language identification on the given recording and then to load the appropriate ASR system. A multilingual system combining the phonetic inventory of several languages into one acoustic model will benefit from total parameter reduction and leaving out the language identification system. Moreover, multilingual system can switch the languages within one utterance. Further research has shown that such multilingual acoustic model also improves all techniques mentioned above [3], [4]. Additionally, these systems can also better handle foreign accented speech [5].

The up-rise of neural network (NN) based features in last few years ported the problem of multilinguality from acoustic modelling to feature extraction process. The NN-based features are obtained from the NN which is trained on data from particular language. Ideally, the same data is used for both, neural network and acoustic model training. In practice, for every acoustic model, also the feature extraction block is trained. The first study of portability of NN-based features was done in [6].

Since then, the NN-based features progressed from the role of cepstral features helper to their competitors so they can be used without cepstral features [7], [8]. This shift makes the question of multilinguality of NN-based features an urgent one – will it be possible to use the NN-based features for new language without the need for (re)training the NN?

Several recent works address the multilinguality of NN: in [9], [10] the authors focused on the usage of NN in language identification system, either through phonotactic model or by training a NN to classify languages. The closest to our problem is the work of Scanzio [11], where a majority of the NN is common for all languages and only the last layer is language specific. This NN, including its language specific part, is then used in a hybrid HMM-ANN.

Our study focuses on the behavior of Probabilistic and Bottle-Neck features [7] trained on a particular language in a system designed for a different language. Further, possibilities of obtaining NN-based features using data from several languages are examined and evaluated.

II. Experimental setup

A. Data

The data comes from multilingual database GlobalPhone [12]. The database covers 19 languages with an average of 20 hours of speech from about 100 native speakers per language. This database aims for an acceptable Out Of Vocabulary (OOV) rate in test sets but with occurrences of words from other languages. This requirement was satisfied by newspaper articles which were read by native speakers. The database covers speakers of both genders in ages from 18 to 81 years. The speech was recorded in office-like environment by high quality equipment. We converted the recordings to 8KHz, 16 bit, mono format.

The following languages were selected for the experiments: Czech (CZ), German (GE), Portuguese (PO), Russian (RU), Spanish (SP), Turkish (TU) and Vietnamese (VN). These languages were accompanied with English (EN) taken from Wall Street Journal database. See Tab. I for detailed numbers of

978-1-4673-0367-5/11/$26.00 ©2011 IEEE

TABLE I
NUMBER OF SPEAKERS AND AMOUNT OF AUDIO MATERIAL IN HOURS OVERALL, FOR TRAINING, DEVELOPMENT AND TESTING

Lang.	Speakers	Audio	TRAIN	DEV	TEST
GE	77	18	13.2	1.8	1.3
CZ	102	29	26.8	1.2	1.9
EN	311	16	14.2	1.0	1.0
SP	100	22	13.4	1.2	1.2
PO	102	26	14.7	1.0	1.0
TU	100	17	12.0	1.6	1.4
VN	129	19	14.7	1.2	1.3
RU	115	22	16.9	1.3	1.4

TABLE II
DETAILED INFORMATION ABOUT LANGUAGE MODELS AND TEST DICTIONARIES FOR INDIVIDUAL TASKS.

Lang	OOV	Dict Size	LM Corpus Size	WWW Server
GE	1.92	375k	19M	www.faz.net
CZ	3.08	323k	7M	www.novinky.cz
EN	2.30	20k	39M	WSJ - LDC2000T43
SP	3.10	135k	18M	www.aldia.cr
PO	0.92	205k	23M	www.linguateca.pt/ cetenfolha
TU	2.60	579k	15M	www.zaman.com.tr
VN	0.02	16k	6M	www.tintuconline.vn
RU	1.44	485k	19M	www.pravda.ru

speakers and data partitioning. Each individual speaker appears only in one set. The partitioning followed the GlobalPhone recommendation.

When preparing the databases, several problems were encountered. The biggest issue was the low quality of dictionaries with many missing words. The Vietnamese dictionary was missing completely. The typos and miss-spelled words were corrected, abbreviations were expanded and missing pronunciations were generated with in-house grapheme-to-phoneme conversion tool. The dictionaries for Vietnamese and Russian were obtained from Lingea[1]. The CMU dictionary was used for English. If the transcription contained a completely unknown word, the sentence was deleted. The final OOV rate is between 1% and 3%, for Vietnamese (syllabic language) it is 0.02%, see Tab. II for details.

The transcription and dictionary encoding was converted to single format suitable for HTK toolkit. Each language has its own phoneme set.

The data for Language Model (LM) were obtained from Internet sources (newspaper articles) using RLAT and SPICE tools[2]. The size of gathered corpus for LM training together with the sources are given in Tab. II. Bigram LMs were generated for all languages except Vietnamese, which is a syllable language and trigram LM was created for it.

[1]http://www.lingea.com

[2]http://i19pc5.ira.uka.de/rlat-dev/index.php, http://plan.is.cs.cmu.edu/Spice/spice/index.php

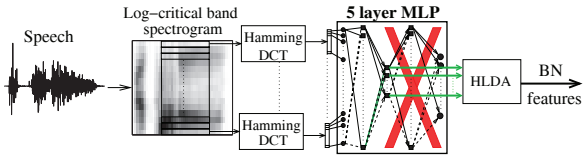

Fig. 1. Block diagram of Bottle-Neck feature extraction

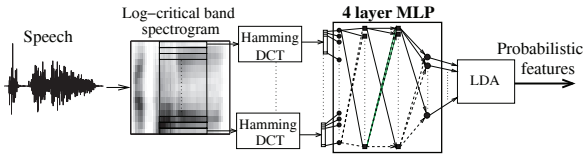

Fig. 2. Block diagram of Probabilistic feature extraction

B. Probabilistic and Bottle-Neck features

The probabilistic and bottle neck features are obtained similarly to [7]: critical band energies are computed from the speech signal first, using standard 25ms window and 10ms shift. A filter-bank with 15 Mel-scaled filters is used. Then, the energies are mean-normalized within the given segments. Further, the block of 31 frames (context of ± 15 frames) is taken and Hamming window followed by Discrete Cosine Transform (DCT) is applied on the trajectory of each energy coefficient. The DCT retrieves 16 coefficients including the 0^{th} one. Finally, there are $15 \times 16 = 240$ coefficients creating input vector for NN training. The TRAIN part of the data is used for NNs training, the DEV part is used as cross-validation set.

The target labels associated with each frame are obtained by forced alignment using a baseline PLP system (see Sec. II-C). The labels represent states of context-independent phoneme models. The numbers of NN targets per language are given in Tab. III.

The Probabilistic and Bottle-Neck features are obtained from a single fully connected NN with approximately 250 000 trainable parameters. This simple approach is preferred over structure of several NNs (although it provides better ASR performance [13]) to keep the experimental setup straight and to avoid expansion of experiments in a direction of NNs structures.

The **Bottle-Neck** (BN) features are obtained from 5-layer NN where the middle hidden layer (BN layer) has the size of the desired feature vector. The first and the third hidden layers have the same size. The outputs of BN layer are decorrelated using Heteroscedastic Linear Discriminant Analysis (HLDA) transform which treats every state of HMM model as a class. Block diagram of BN feature extraction is shown in Fig. 1.
The **Probabilistic** features are obtained by a 4-layer NN where the first and the second hidden layers have the same size. The output probability estimates are processed by logarithm non-linearity and decorrelated by Linear Discriminant Analysis (LDA) which also reduces the vector dimensionality to 30 coefficients. LDA classes corresponds to NN targets. The use of HLDA was not possible because of insufficient data to collect the statistics (large input vector) for the transformation

TABLE III
NUMBER OF NN TARGETS AND BASELINE RESULTS [WER%] FOR SYSTEMS TRAINED ON PLP, PROBABILISTIC AND BN FEATURES

Lang.	NN targets	results with features		
		PLP	Prob	BN
GE	129	28.2	26.8	25.4
CZ	129	24.2	20.4	20.1
EN	123	17.6	17.4	16.2
SP	126	25.0	23.4	24.1
PO	141	28.0	26.5	26.1
TU	93	34.5	31.6	31.6
VN	258	28.4	30.5	23.5
RU	174	35.4	32.9	33.6

computation and resulting features provided worse recognition performance then those with LDA transform. Block diagram of Probabilistic feature extraction is given in Fig. 2.

Since this study is focused on the behaviour of language-dependent NN-based features in systems built for other languages, the resulting features are not augmented with any other features (PLP, deltas).

C. Recognition system

The recognizer system is based on HMM cross-word tied-states triphones acoustic models. The models contain ≈3000 tied states with 18 Gaussian mixtures per state. Models for each parameter set were trained from scratch using mixture-up maximum likelihood training.

Mel-filter bank based PLP coefficients were used as language independent parameters. There were 13 direct parameters augmented with deltas and double-deltas totaling in feature vectors with 39 coefficients. Cepstral mean and variance normalization was applied on speaker basis. The resulting models were used for forced alignment of the data preceding the NN training.

Apart from the PLP baseline, the 4- and 5-layer NNs were trained on each language representing the ideal case of the same data for NN and acoustic model training. All these results are given in Tab. III. It can be seen that for all languages except Vietnamese Probabilistic features, the NN-based features performs better then the PLP features. The poor results obtained by Probabilistic features in case of Vietnamese can be explained by large number of classes and inefficient dimensionality reduction. As can be seen, this problem is overcome by the dimensionality reduction embedded in NN architecture in the BN feature extraction.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Since the study is focused on the behavior of NN-based features obtained on one language in different language system, the division between "*old*" and "*new*" languages has to be made. The *old* languages should represent well established systems and databases for their training: we have selected German, Czech and English. This languages are chosen for our previous experience with them and also for the availability of native speakers in our group who are able to check the correctness of transcripts and dictionary entries. The rest of the languages will play the role of *new* languages.

TABLE IV
RESULTS [WER%] OF NN BASED FEATURES FROM *Old* LANGUAGES APPLIED TO *New* LANGUAGES.

Lang.	NN trained on					
	GE		CZ		EN	
	Prob	BN	Prob	BN	Prob	BN
SP	24.9	30.6	24.9	26.8	27.1	29.3
PO	28.4	31.5	27.6	27.9	29.7	31.3
TU	37.5	39.9	34.1	35.0	36.0	37.6
VN	34.7	37.8	30.9	32.7	34.1	34.9
RU	37.5	40.7	33.2	35.9	37.7	39.8

A. Applying NN in different language

The first step in our evaluation is to simply apply *old* NN-based features into a *new* language. The results are given in Tab. IV. This experiment would represent the worst scenario in the sense, that only a single *old* language is available. Note, that the decorrelation step is done on the *new* language data.

By taking a look at Tab. IV, it can be seen that from BN features, the best performing are the CZ ones. Comparing them with Tab. III, an increase of WER is observed, but for some languages (PO, TU, RU) the features are still competitive to language-independent PLP features. Although the Probabilistic features are inferior to BN features in matched language scenario, in this case they are superior. We hypothesize that this behavior is caused by the larger variability in the probability outputs and that the following LDA decorrelation is able to find the directions important for the new language.

B. Concatenation of NN outputs

This scenario is aimed at a fast application of already trained NNs into new language domain. The simple way, to combine outputs of several NN is to merge their outputs. Several methods of probability averaging were proposed for multistream approach (for example [14]) but this technique assumes the same target classes for all NNs. In this scenario, each NN is trained on different language and thus having different target classes. Moreover, the averaging approach cannot be used for BN outputs. On the other hand, our goal is not to precisely classify given classes, but to obtain features for the subsequent model. Since the decorrelation step is always in the processing chain, it can also effectively handle the task of NN output merging.

The *new* language data are processed by two (three,...) NNs in parallel and the output vectors from these NNs are concatenated. The resulting vector is decorrelated and desired dimensionality obtained. The results achieved by concatenation of (e.g. CZ and GE ⇒ CZ.GE) NNs are given in Tab. V. Looking at the results obtained with BN features, it can be seen that they are similar to the one obtained with Czech NN only. This indicates, that improving resulting BN features by the means of parallel data processing by several NNs might be more difficult. The positive finding is that concatenated systems are more robust and the system performance is close to the best one achieved by a single *old* language NN (which is not known beforehand). The probabilistic features performs better again. All the systems have almost the same performance indicating

TABLE V
RESULTS [WER%] OF FEATURES DERIVED FROM CONCATENATED OUTPUTS OF *Old* LANGUAGES NNS APPLIED TO *New* LANGUAGES.

	NN trained on					
	CZ.GE		CZ.EN		CZ.GE.EN	
Lang.	Prob	BN	Prob	BN	Prob	BN
SP	25.6	27.3	25.2	26.9	25.8	27.0
PO	27.9	28.6	27.8	28.5	28.5	28.0
TU	34.9	35.1	34.1	34.7	35.1	35.3
VN	30.9	32.7	30.6	31.4	31.0	32.0
RU	33.6	35.9	33.6	35.4	33.6	35.8

TABLE VI
LIST OF PHONEME LABELS SHARING THE SAME IPA SYMBOL AND EXAMPLE WORDS.

GE	CZ	examples	GE	CZ	examples
b	b	**B**all, **b**yl	S	sh	**sch**al, šelest
d	d	**d**ann, **d**élka	t	t	**T**al, **t**en
f	f	**F**ass, **f**oukat	ts	c	**Z**ahl, **c**ena
g	g	**G**ast, **g**ag	v	v	**w**as, **v**ítr
h	h	**h**at, **h**ořet	x	x	Ba**ch**, **ch**omout
j	j	**j**a, **j**enom	z	z	Ha**s**e, **z**ima
k	k	**k**alt, **k**olo	a	a	D**a**ch, m**a**tka
l	l	**L**ast, **l**ak	al	aa	B**ah**n, m**á**ma
m	m	**M**ast, **m**ouka	e	e	B**e**tt, l**e**t
n	n	**N**aht, **n**yn	el	ee	w**äh**le, l**é**tat
ng	ng	la**ng**, Ha**n**ka	il	ii	v**ie**l, kl**í**t
p	p	**P**akt, **p**yl	i	i	b**i**st, kl**i**d
r	r	**R**ast, **r**obot	ol	oo	B**oo**t, m**ó**da
s	s	Ha**s**t, **s**tůl	ul	uu	H**u**t, k**ů**lna

that the information can be efficiently utilized and systems created in this way are stable.

C. Phoneme set unification

The next step is to train a single NN covering targets from several languages. Similarly to the creation of multilingual acoustic models, data-driven or knowledge-based approaches can be used. It is also possible to join the target sets of individual languages into one. To avoid expansion of number of experiments, only CZ and GE are considered in this part.

The solution for the latter mentioned case is straightforward: all labels except silence are appended with language identification mark. The data are then mixed (within the given set) and one single NN is trained.

The knowledge base unification of phoneme sets is done through IPA notation: First, the phonemes labels are mapped to IPA symbols using the example words given in IPA and their transcription in dictionary. Then the labels sharing the same IPA symbols are assigned a new label. The rest of phonemes stay language-dependent. The merged phoneme labels with example words in both languages can be seen in Tab. VI.

The data driven approach was based on the phoneme hard confusion matrix. The Czech NN was used to classify German data. The value of the hard confusion at position i, j: $H_{i,j}$ is given by the number of times when the input vector with label GE_i is classified into the output class CZ_j. Then, the normalized matrix is computed in the following way: $Hn_{i,j} = \frac{H_{i,j}}{\sum_{\forall j} H_{i,j}}$, where $\sum_{\forall j} H_{i,j}$ is the number of vectors belonging to input label GE_i. The labels GE_i and CZ_j are said to be the same if $Hn_{i,j} > 0.4$. This threshold was set experimentally to allow only one GE_i to be mapped to CZ_j and to get the maximum number of merged labels at the same time.

TABLE VII
RESULTS [WER%] OF NN BASED FEATURES FROM NNS WITH UNIFIED PHONEME TARGETS.

	Unification method					
	join		IPA		Data driven	
Lang.	Prob	BN	Prob	BN	Prob	BN
SP	26.8*	26.9	25.2	27.1	26.0	27.1
PO	28.7*	27.9	27.9	28.6	28.1	28.1
TU	36.7	35.2	34.4	35.3	34.2	35.7
VN	33.7	32.4	31.7	33.4	31.5	34.5
RU	36.3*	34.9	33.7	35.7	33.5	35.6

* variance of one coefficient was found to be 0.

The results obtained with different unification scenarios are given in Tab VII. Comparing the BN feature results with one another, it can be observed that simple joining of phoneme sets yields slightly better results then the other methods. But this method fails for Probabilistic features. The decorrelation transform was badly estimated which resulted in zero variance of one coefficient in resulting feature vector (look for * mark in Tab. VII). The problem was solved by flooring the variance at 10^{-5} during GMM training. The system ended up with worse performance then other unification method.

The knowledge based and data driven unification methods provides better results with Probabilistic features, but within the given features, both kids of unification give about the same results.

Note the interesting behavior of BN features: the system based on join phoneme set is able to preserve the phoneme variability between languages although the BN layer function as a compression and (almost) the same phoneme classes should produce the same output. But when the (almost) same phonemes are merged into one, this source of variability is lost. Thus NN cannot learn the slight differences useful in languages with different phoneme sets. This is illustrated by the increase of WER for systems with unified phoneme set.

D. NN trained on large data

Since the data in our database is quite small (about 15 hours for training) we were interested what will be the performance with NN trained on larger training set. For this purpose, a part of the SwitchBoard corpus was utilized. This data is Conversation Telephone Speech (CTS) which is different speaking style from read speech we have been using so far. Additional differences are caused by the technical parameters of the recordings: telephone channel causes band limitation of the recordings and adds noises to speech signal. This differences are another subject of interest - will the amount of training data have positive effect on the system performance or will it be outweighed by the difference between train and test data?

100 hours of speech were randomly selected from SwitchBoard and NNs with approximately 1 000 000 parameters were

TABLE VIII
RESULTS [WER%] OF NN BASED FEATURES FROM CTS NNS.

	EN CTS features					
	alone		CZ concat		CZGE concat	
Lang.	Prob	BN	Prob	BN	Prob	BN
EN	15.1	15.6				
SP	26.1	25.6	24.1	25.6	24.6	26.4
PO	28.5	28.5	26.4	28.1	26.8	28.4
TU	33.5	32.1	32.3	32.6	32.8	34.0
VN	31.6	29.1	28.7	28.5	29.4	28.7
RU	35.4	35.1	32.6	34.2	33.1	34.8

trained. Since the phoneme alignment of the CTS data is also based on the CMU dictionary, the phoneme set is the same as for our English task. The performance of the EN CTS features on GlobalPhone data is given on the first line in Tab. VIII. This shows, that using larger dataset which allows to train bigger NNs can be beneficial for cross-domain databases (conversations vs. read speech).

The performance of features obtained through EN CTS NNs on other languages are given in the first two columns in Tab. VIII. Comparing it with results obtained by English NNs in Tab. IV, considerable and consistent improvement can be seen. Moreover, these are the best performing single language features.

Concatenating these features to Czech ones (see 3^{rd} and 4^{th} columns in Tab. VIII), an improvement is achieved in most cases. Further concatenation with German features (last two columns) brings impairment of the resulting system.

IV. CONCLUSION

This study focused on the performance of NN-based features in multilingual environment. Since the NN is usually trained on the same data as the acoustic model, porting already trained NN to a new language might be problematic. In our scenario, we decided to have three *old* languages, for which the NNs are already trained, and five *new* languages, where these NNs were applied.

It has been shown, that NN-based features can perform well when applied on different languages and they can still outperform language-independent PLP features. An important observation is that the Probabilistic features perform better than BN features when applied on different language. Our hypothesis is that the Probabilistic features contain larger variability which can be utilized in the decorrelation and dimensionality reduction steps which is done with respect to target language.

Further experiments with concatenating of several NNs outputs reveal that the resulting features have performance close to the best individual ones. This is especially the case for probabilistic features which reach practically the same performance. This is an important finding as in practice, we usually cannot test the systems on new data.

The experiment with unifying the phoneme set confirmed our hypothesis that the variability in output vector is important, because with compacting the phoneme set, the performance decreased. It was also observed that simple concatenation of NNs outputs brought similar performance as NN newly trained on several languages. This is again important finding showing that it is possible to simply concatenate outputs of existing networks without the necessity of training a new one on unified languages.

Finally, we have performed experiments with NN trained on large amount of Conversation Telephone Speech which differ in speaking style and recording environment from our test data. This experiment shows that the large amount of data from different domain is useful not only for the same language tasks, but can be efficiently used also for other languages. The resulting features outperformed not only the original English ones, but also all systems trained before. Concatenating these features with Czech ones (the best features obtained on GlobalPhone data) slightly improves the resulting performance and gives better or the same results compared to language-independent PLP features.

This result suggests, that there is the possibility to further improve NN-based feature extraction for unseen language. Our future work will focus on the variability issue and NNs with larger BN will be trained to allow the dimensionality reduction to find useful directions for new language.

ACKNOWLEDGMENT

This work was partly supported by Czech Ministry of Trade and Commerce project No. FR-TI1/034. F. Grézl was supported by Grant Agency of Czech Republic post-doctoral project No. GP102/09/P635. The work was also supported by Czech Ministry of Education project No. MSM0021630528 and Grant Agency of Czech Republic project No. 102/08/0707.

REFERENCES

[1] A. Constantinescu and G. Chollet, "On cross-language experiments and data-driven units for alisp (automatic language independent speech processing)," in *Proc. ASRU 1997*, dec 1997, pp. 606 –613.

[2] B. Wheatley, K. Kondo, W. Anderson, and Y. Muthusamy, "An evaluation of cross-language adaptation for rapid hmm development in a new language," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 1, pp. 237–240, 1994.

[3] U. Bub, J. Kohler, and B. Imperl, "In-service adaptation of multilingual hidden-markov-models," in *Proc. ICASSP 1997*. IEEE Signal Processing Society, 1997, pp. 1451–1454.

[4] J. Köhler, "Language adaptation of multilingual phone models for vocabulary independent speech recognition tasks," in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, vol. 1, may 1998, pp. 417 –420 vol.1.

[5] S. Witt and S. Young, "Language learning based on non-native speech recognition," in *In Proceedings of Eurospeech*, 1997, pp. 633–636.

[6] A. Stolcke, F. Grézl, M. Hwang, X. Lei, N. Morgan, and D. Vergyri, "Cross-domain and cross-language portability of acoustic features estimated by multilayer perceptrons," in *Proceedings of ICASSP 2006*, Toulouse, FR, 2006, pp. 321–324.

[7] F. Grézl, M. Karafiát, S. Kontár, and J. Černocký, "Probabilistic and bottle-neck features for LVCSR of meetings," in *Proc. ICASSP 2007*, Honolulu, Hawaii, USA, Apr 2007, pp. 757–760.

[8] F. Grézl and P. Fousek, "Optimizing bottle-neck features for LVCSR," in *2008 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 4729–4732.

[9] D. Imseng, M. Magimai.-Doss, and H. Bourlard, "Hierarchical multilayer perceptron based language identification," in *Proceedings of Interspeech*, sep 2010, pp. 2722–2725.

[10] A. Stolcke, M. Akbacak, L. Ferrer, S. Kajarekar, C. Richey, N. Scheffer, and E. Shriberg, "Improving language recognition with multilingual phone recognition and speaker adaptation transforms," in *Proceedings of the Odyssey Speaker and Language Recognition Workshop*, Jun. 2010, pp. 256–262.

[11] S. Scanzio, P. Laface, L. Fissore, R. Gemello, and F. Mana, "On the use of a multilingual neural network front-end," in *Proceedings of INTERSPEECH-2008*, 2008, pp. 2711–2714.

[12] T. Schultz, M. Westphal, and A. Waibel, "The globalphone project: Multilingual lvcsr with janus-3," in *in Multilingual Information Retrieval Dialogs: 2nd SQEL Workshop, Plzen, Czech Republic*, 1997, pp. 20–27.

[13] F. Grézl, M. Karafiát, and L. Burget, "Investigation into bottle-neck features for meeting speech recognition," in *Proc. Interspeech 2009*, Sep 2009, pp. 294–2950.

[14] H. Misra, H. Bourlard, and V. Tyagi, "New entropy based combination rules in hmm/ann multi-stream asr," in *IN PROC. ICASSP 2003, HONG KONG*, 2003, pp. 741–744.