The impact of accent identification errors on speech recognition of South African English

AUTHORS:
Herman Kamper[1]
Thomas R. Niesler[1]

AFFILIATIONS:
[1]Department of Electrical & Electronic Engineering, Stellenbosch University, Stellenbosch, South Africa

CORRESPONDENCE TO:
Thomas Niesler

EMAIL:
trn@sun.ac.za

POSTAL ADDRESS:
Department of Electrical & Electronic Engineering, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

DATES:
Received: 26 Oct. 2012
Revised: 06 Sep. 2013
Accepted: 09 Oct. 2013

KEYWORDS:
parallel recognition; acoustic modelling; human language technology

HOW TO CITE:
Kamper H, Niesler TR. The impact of accent identification errors on speech recognition of South African English. S Afr J Sci. 2014;110(1/2), Art. #2012-0049, 6 pages. http://dx.doi.org/10.1590/sajs.2014/20120049

For successful deployment, a South African English speech recognition system must be capable of processing the prevalent accents in this variety of English. Previous work dealing with the different accents of South African English has considered the case in which the accent of the input speech is known. Here we focus on the practical scenario in which the accent of the input speech is unknown and accent identification must occur at recognition time. By means of a set of contrastive experiments, we determine the effect which errors in the identification of the accent have on speech recognition performance. We focus on the specific configuration in which a set of accent-specific speech recognisers operate in parallel, thereby delivering both a recognition hypothesis as well as an identified accent in a single step. We find that, despite their considerable number, the accent identification errors do not lead to degraded speech recognition performance. We conclude that, for our South African English data, there is no benefit of including a more complex explicit accent identification component in the overall speech recognition system.

Introduction

Although English is used throughout South Africa, it is spoken as a first language by just 9.6% of the population.[1] Accented English is therefore highly prevalent and speech recognition systems must be robust to these different accents before successful speech-based services can become accessible to the wider population.

One solution to accent-robust automatic speech recognition is to develop acoustic models that are accent independent. For accent-independent models, no distinction is made between different accents, and training data are pooled. A different approach is to develop several separate acoustic model sets that are each designed to deliver optimal performance for a particular accent and then to combine these within a single system. In this latter case, accent identification (AID) must occur in order for the correct set of acoustic models to be chosen at recognition time. However, state-of-the-art AID is complex and adds a significant additional hurdle to the development of a speech recognition system.[2] In this study we investigated the impact that such AID errors have on the accuracy of speech recognition for the five acknowledged accents of South African English (SAE).To do this, we compared three established acoustic modelling approaches by means of a set of contrastive speech recognition experiments in which the accent of the SAE input speech is assumed to be unknown. As a baseline, we also considered the performance that was achieved when the input accent is known. Our results provide some insight into the importance of AID accuracy within the context of an SAE speech recognition system.

Accents of English in South Africa

Five major accents of SAE are identified in the literature: Afrikaans English, Black South African English, Cape Flats English, White South African English and Indian South African English.[3] The term 'South African English' is used to refer collectively to all the accents of English spoken in the country.

English itself was originally brought to South Africa by British occupying forces at the end of the 18th century. The accent known as White South African English (EE) refers to the first-language English spoken by White South Africans who are chiefly of British descent. This accent is used by approximately 3.1% of the population.[1] Afrikaans English (AE) refers to the accent used by White South African second-language English speakers of Afrikaans descent. Afrikaans has its origins in 17th century Dutch, which was brought to South Africa by settlers from the Netherlands. Although the Afrikaans vocabulary has a predominantly Dutch origin, it was influenced by Malay, Portuguese and the Bantu and Khoisan languages. White Afrikaans speakers comprise approximately 5.3% of the South African population.[1] Cape Flats English (CE) has its roots in the 19th century working-class residential areas in inner-city Cape Town where residents from many different ethnic affiliations, religions and languages came into regular contact with one another. The accent spread as residents from these mixed neighbourhoods moved or were forced to move to the Cape Flats in the 1960s and 1970s.[4] Today CE is most closely associated with the 'Coloured' community of South Africa, which comprises approximately 9.1% of the total population.[1] The connection between EE, AE and CE, all three of which have been strongly influenced by Afrikaans, has been emphasised in the literature.[3] Black South African English (BE) refers to the variety of English spoken by Black South Africans whose first language is an indigenous African language. BE speakers are overwhelmingly first-language speakers of one of the nine official indigenous African languages of South Africa and comprise approximately 74.9% of the population.[1] Indian languages were brought to South Africa by labourers who were recruited from India after the abolition of slavery in European colonies in the 19th century. These Indian languages have existed in South Africa since 1860, mainly in KwaZulu-Natal. Today Indian South African English (IE) is spoken as a first language by most of the Indian South African population. Approximately 2.5% of the South African population are considered Indian or Asian and 86.1% speak English as a first language.[1]

To obtain some initial intuition regarding the relative similarity of these five accents, we determined how similar the statistical distributions describing the acoustics of corresponding sounds in each accent were to one another. We achieved this by applying the Bhattacharyya distance, which allows a measure of similarity between two probability density functions to be computed.[5] Three-state single-mixture monophone hidden Markov models (HMMs) were obtained using the acoustic data for each accent separately. For each accent pair, the Bhattacharyya

distance was subsequently computed between corresponding states of the two HMMs. The average over the three resulting distances was then determined to obtain a measure of between-accent similarity for a particular monophone. Finally, the average distance between all corresponding pairs of monophones was calculated to obtain a measure of inter-accent similarity.[6] For the five accents of SAE, an approximate representation of these distances is shown in Figure 1, where particle swarm optimisation[7] was used to find an approximate projection of the distances into two dimensions. In the figure, similarity is indicated by a geometrically shorter distance between accents. From this first analysis we conclude that AE, CE and EE are quite similar, while BE and IE are more dissimilar from the other accents and from each other.

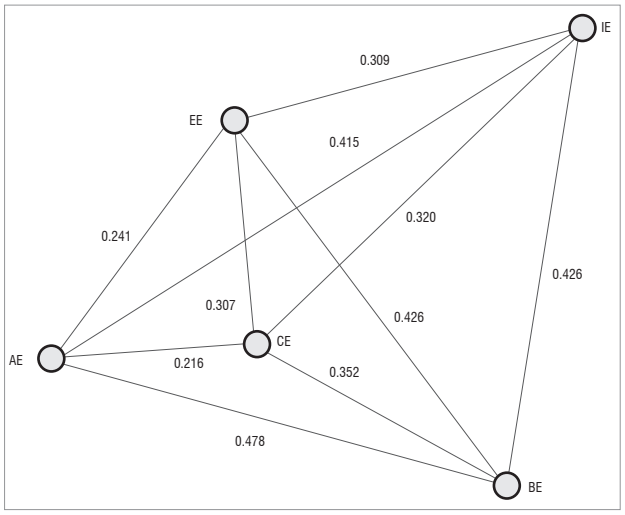

Figure 1: Graphical depiction of the average Bhattacharyya distances between the five accents of South African English: White South African English (EE), Indian South African English (IE), Black South African English (BE), Afrikaans English (AE) and Cape Flats English (CE).

Related research

Several studies have considered acoustic modelling for different accents of the same language. Approaches include the pooling of data across accents, leading to a single accent-independent acoustic model[8]; the isolation of data for each accent, leading to individual accent-specific acoustic models[9]; and adaptation techniques in which models trained on one accent are adapted using data from another[10,11]. Recently, selective data sharing across accents through the use of appropriate decision-tree state clustering algorithms has also received some attention.[6,12] These studies extend the multilingual acoustic modelling approach first proposed by Schultz and Waibel[13] to apply to multiple accents of the same language.

Most of the above studies consider the scenario in which the accent of the incoming speech is known and each utterance is presented only to the matching set of acoustic models. This approach is appropriate when the aim is to evaluate different acoustic modelling strategies without allowing performance to be influenced by the effects of accent misclassification. Because the accent is assumed to be known, we will refer to this approach as oracle AID. However, in many practical situations, the accent of the incoming speech would not be known. In such cases a single system should be able to process multiple accents.

Three approaches for the recognition of multiple accents are commonly found in the literature. One approach is to precede accent-specific speech recognition with an explicit AID step.[14] A second approach is to run a bank of accent-specific recognisers in parallel and select the output with the highest associated likelihood.[15] In this set-up, AID is performed implicitly during recognition. A third approach is to train a single accent-independent model set by pooling data across accents and thereby avoid AID altogether.[15]

These three approaches have been applied in various ways to different English accents. The recognition of non-native English from six European countries was considered by Teixeira et al.[16] They found that AID followed by recognition gave comparable performance to an oracle configuration, but both were outperformed by an accent-independent system. Chengalvarayan[15] compared the parallel and accent-independent approaches for recognition of American, Australian and British English and found that the accent-independent approach gave the best performance. Variations of the three approaches have also been considered. For example, Beattie et al.[17] proposed a parallel recognition approach in which the accent-specific recogniser is selected based on a history of scores for a specific speaker instead of only the score for the current utterance. This method proved to be superior to accent-independent modelling for three dialects of American English. We also compared the oracle, parallel and accent-independent strategies in our experiments to determine what could be learnt for the case of the accents of SAE.

Speech resources

Training and test data

Our experiments were based on the African Speech Technology (AST) databases.[18] The databases consist of telephone speech recorded over fixed and mobile telephone networks and contain a mix of read and spontaneous speech. As part of the AST project, five English accented speech databases were compiled corresponding to the five accents of SAE. These databases were transcribed both phonetically, using a common international phonetic alphabet (IPA)-based phone set consisting of 50 phones, as well as orthographically. The assignment of a speaker's English accent was guided by the speaker's first language and ethnicity.

Each of the five databases was divided into training, development and evaluation sets. As indicated in Tables 1 and 2, the training sets each contain between 5.5 h and 7 h of speech from approximately 250 speakers, while the evaluation sets contain approximately 25 min from 20 speakers for each accent. The development sets were used only for the optimisation of the recognition parameters before final testing on the evaluation data. For the development and evaluation sets, the ratio of male to female speakers is approximately equal and all sets contain utterances from both landline and mobile telephones. There is no speaker overlap between any of the sets. The average length of a test utterance is approximately 2 s.

Table 1: Training sets for each South African English accent

Accent	Number of hours of speech	Number of utterances	Number of speakers	Word tokens
AE	7.02	11 344	276	52 540
BE	5.45	7779	193	37 807
CE	6.15	10 004	231	46 185
EE	5.95	9879	245	47 279
IE	7.21	15 073	295	57 253
Total	31.78	54 078	1240	241 064

AE, Afrikaans English; BE, Black South African English; CE, Cape Flats English; EE, White South African English; IE, Indian South African English.

Table 2: Evaluation sets for each South African English accent

Accent	Speech (min)	Number of utterances	Number of speakers	Word tokens
AE	24.16	689	21	2913
BE	25.77	745	20	3100
CE	23.83	709	20	3073
EE	23.96	702	18	3059
IE	25.41	865	20	3362
Total	123.13	3710	99	15 507

AE, Afrikaans English; BE, Black South African English; CE, Cape Flats English; EE, White South African English; IE, Indian South African English.

Language models and pronunciation dictionaries

Using the SRI language modelling (SRILM) toolkit,[19] an accent-independent backoff[20] bigram language model was trained on the combined training set transcriptions of all five accents. Absolute discounting was used for the estimation of language model probabilities.[21] As part of the AST project, a separate pronunciation dictionary was obtained for each accent individually. These individual contributions were combined into a single pronunciation dictionary for the experiments presented here. These design decisions were made based on preliminary experiments which indicated that accent-independent pronunciation and language modelling outperformed the accent-specific alternatives. Language model perplexities and out-of-vocabulary rates are shown in Table 3.

Table 3: Bigram language model perplexities and out-of-vocabulary (OOV) rates measured on the evaluation sets

Accent	Bigram types	Perplexity	OOV (%)
AE	11 580	24.07	1.82
BE	9639	27.87	2.84
CE	10 641	27.45	1.40
EE	10 451	24.90	1.08
IE	11 677	25.55	1.73

AE, Afrikaans English; BE, Black South African English; CE, Cape Flats English; EE, White South African English; IE, Indian South African English.

Experimental methodology

General set-up

Speech recognition systems were developed using the HTK tools.[22] Speech audio data were parameterised as 13 Mel-frequency cepstral coefficients with their first- and second-order derivatives to obtain 39-dimensional observation vectors. Cepstral mean normalisation was applied on a per-utterance basis. The parametrised training sets were used to obtain three-state left-to-right single-mixture monophone HMMs with diagonal covariance matrices using embedded Baum-Welch re-estimation. These monophone models were then cloned and re-estimated to obtain initial cross-word triphone models which were subsequently subjected to decision-tree state clustering. Clustering was followed by five iterations of re-estimation. Finally, the number of Gaussian mixtures per state was gradually increased, with each increase being followed by five iterations of re-estimation. This yielded diagonal-covariance cross-word tied-state triphone HMMs with three states per model and eight Gaussian mixtures per state.

Acoustic modelling

When performing speech recognition of multiple accents by running separate recognisers in parallel, different acoustic modelling approaches can be followed. We considered two modelling approaches. Comparable acoustic modelling approaches have been previously considered in multilingual[13,23] as well as multi-accent[6,12,24] settings. The two approaches we used – accent-specific acoustic modelling and multi-accent acoustic modelling – are distinguished by different methods of decision-tree state clustering.

In accent-specific acoustic modelling, separate accent-specific acoustic models are trained and no sharing of data occurs between accents. Separate decision-trees are grown for each accent and the clustering process employs only questions relating to phonetic context. Chengalvarayan[15] has applied such models in a parallel configuration. In multi-accent acoustic modelling, decision-tree questions relate not only to the phonetic context, but also to the accent of the basephone. Tying across accents can thus occur when triphone states are similar, while the same triphone state from different accents can be modelled separately when there are differences. Detailed descriptions of this approach can be found in the existing literature.[6,23]

As a further benchmark we considered accent-independent acoustic modelling in which a single model set is obtained by pooling data across all accents for phones with the same IPA classification and the need for AID is side-stepped. The decision-tree clustering process employs only questions relating to phonetic context. Such pooled models are often employed for the recognition of accented speech[12,15,16] and therefore represent an important baseline.

System configuration and evaluation

Five recognisers, one tailored to each SAE accent, were configured in parallel and the output with the highest associated likelihood was selected as the final speech recognition hypothesis. This configuration is illustrated in Figure 2. The selection of the highest scoring result can be performed independently for each utterance, leading to per-utterance AID, or for each speaker, leading to per-speaker AID. The choice between these two AID schemes will depend on practical constraints and we report on the performance of both. Accent-specific and multi-accent acoustic models were employed in competing parallel systems and these were compared with accent-independent acoustic models. As a further benchmark we compared the performance of the parallel systems to that of systems in which each test utterance was presented only to the recogniser with matching accent (oracle AID). This configuration is illustrated in Figure 3. In each case the oracle configuration used the same acoustic models as the parallel configuration it was compared to. In this way the penalty as a result of AID errors occurring implicitly during parallel recognition can be analysed by direct comparison with the oracle configuration.

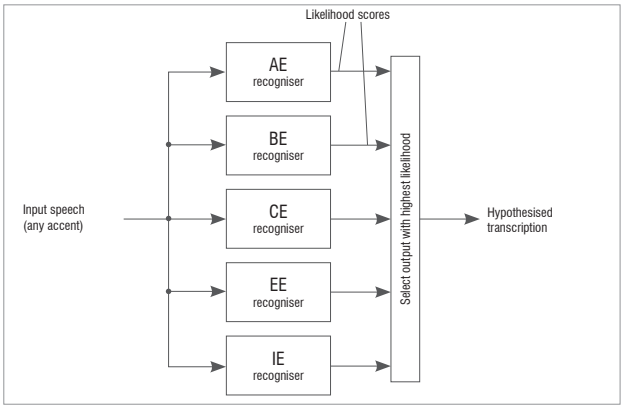

Figure 2: Parallel configuration in which multiple accent-specific recognisers are placed in parallel for simultaneous speech recognition in the five South African English accents: White South African English (EE), Indian South African English (IE), Black South African English (BE), Afrikaans English (AE) and Cape Flats English (CE).

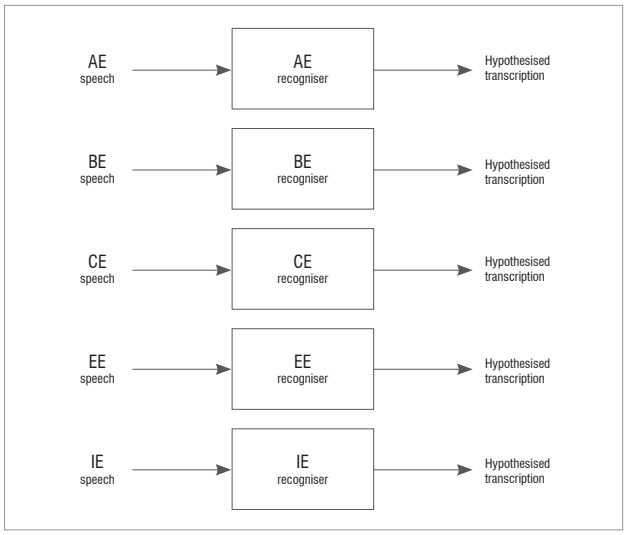

Figure 3: Oracle configuration in which each test utterance is presented only to the accent-specific recognition system with matching accent for speech recognition in the five South African English accents: White South African English (EE), Indian South African English (IE), Black South African English (BE), Afrikaans English (AE) and Cape Flats English (CE).

Experimental results

Table 4 shows the average word recognition and per-utterance AID accuracies measured on the evaluation sets. Implicit per-utterance AID was performed by the parallel accent-specific and multi-accent systems. Because a single recogniser was used for the accent-independent system, AID did not occur and identical results were obtained for the oracle and parallel configurations.

Table 4: Recognition performance of the oracle and parallel configurations when applying per-utterance accent identification (AID). Average word recognition accuracies (%) are given for oracle and parallel configurations and per-utterance accuracies (%) are given for AID.

Model set	Oracle	Parallel	AID
Accent-specific	81.53	81.31	67.60
Accent-independent	81.67	81.67	–
Multi-accent	82.78	82.85	65.39

The results in Table 4 indicate that the parallel configuration employing accent-specific acoustic models (with an accuracy of 81.31%) was outperformed by its corresponding oracle configuration (with an accuracy of 81.53%). In contrast, the parallel configuration employing multi-accent acoustic models (82.85%) showed a small improvement over its oracle counterpart (82.78%). This improvement was despite the fact that the accent was misclassified in 34.6% of test cases. Although the improvement in recognition performance was small and only significant at the 66% level,[25] it is noteworthy that the accent misclassifications did not lead to deteriorated accuracy. This observation indicates that some test utterances may have been better matched to the acoustic models of another accent. The results in Table 4 also confirm earlier reports in the literature of the better performance achieved by multi-accent acoustic models in relation to their accent-specific and accent-independent counterparts.[6] It is also apparent that, despite their better speech recognition performance, the multi-accent models did not lead to improved AID.

The performance of parallel configurations applying per-speaker AID is shown in Table 5. The performance of the oracle configurations and the accent-independent model set were unchanged from Table 4. A comparison between the two tables indicates that performing AID on a per-speaker basis improved speech recognition accuracy (from 81.31% to 81.69%) for the parallel systems using accent-specific acoustic models, while recognition accuracy (82.85%) was unchanged for the multi-accent systems. In both cases, AID accuracies were substantially improved to approximately 90%. Among the acoustic modelling options, the multi-accent approach continued to deliver the best performance.

Table 5: Recognition performance of the oracle and parallel configurations when applying per-speaker accent identification (AID). Average word recognition accuracies (%) are given for oracle and parallel configurations and per-utterance accuracies (%) are given for AID.

Model set	Oracle	Parallel	AID
Accent-specific	81.53	81.69	89.84
Accent-independent	81.67	81.67	–
Multi-accent	82.78	82.85	89.81

The results presented here lead us to the surprising conclusion that better AID does not necessarily lead to higher speech recognition accuracy for the accents of SAE in our databases. Despite a sizeable proportion of accent misclassifications (34.6%), the parallel per-utterance AID system employing multi-accent acoustic models showed no deterioration in accuracy compared to an oracle configuration (in which no accent misclassifications occur). When parallel recognition was performed at a per-speaker level, the proportion of AID errors reduced substantially from 34.6% to 10.2%, but the performance of the parallel multi-accent system remained unchanged.

A reason for this weak dependency of speech recognition accuracy on AID performance could be the inherent difficulty of classifying accents. Accent labels were assigned on the basis of the speaker's first language and ethnicity, which may not be a reliable way to determine 'true' accent. Recent research has shown, for example, that Black speakers with an indigenous African language as a first language exhibit an increasing tendency of adopting the accent normally attributed to White South African speakers of English.[26] Furthermore, even when the accent label is uncontested, there may be considerable variation.[27,28] Hence the accent labels may in some cases not be ideal from the point of view of acoustic homogeneity.

Analysis of accent misclassifications

An accent misclassification occurs when the accent of the recogniser selected for a particular utterance during parallel recognition is different from the accent with which that utterance is labelled. We analysed these errors for the case of per-utterance AID. For each misclassified utterance, the recognition hypothesis produced by the oracle and the parallel configurations were obtained. Both these hypotheses were subsequently aligned with the reference transcription. By comparing the two alignments we determined whether the misclassification resulted in an improvement or in a deterioration in performance relative to that of the oracle configuration. The resulting effect on recognition performance of using the parallel configuration instead of the oracle configuration was also calculated.

Tables 6 and 7 present the results of this analysis for the accent-specific and multi-accent acoustic model sets, respectively. We see that, for both acoustic modelling approaches, the majority (approximately 75%) of misclassified utterances did not influence speech recognition performance. While misclassified utterances were shorter than the average evaluation set utterance (~1.7 s compared with ~2 s), those which led to improved performance were longer (~2.7 s). Misclassified utterances leading to deteriorated accuracy were of approximately average length (~2.1 s) while utterances having no effect were shorter (~1.4 s). These observations apply to both acoustic modelling approaches.

Table 6: Analysis of accent misclassifications for the parallel per-utterance accent identification system using accent-specific acoustic models

Impact of misclassification accuracy	Number of utterances	Number of tokens	Average duration (s)	Δ accuracy (%)
No effect	905	2721	1.42	0
Improvement	135	802	2.66	+1.29
Deterioration	162	773	2.22	−1.52
Total/Average†	1202	4296	1.67†	−0.23

Table 7: Analysis of accent misclassifications for the parallel per-utterance accent identification system using multi-accent acoustic models

Impact of misclassification accuracy	Number of utterances	Number of tokens	Average duration (s)	Δ accuracy (%)
No effect	1040	3213	1.46	0
Improvement	120	705	2.66	+1.19
Deterioration	124	559	2.05	−1.12
Total/Average†	1284	4477	1.63†	+0.07

Focusing first on the accent-specific systems, Table 4 indicates that the parallel configuration (81.31% accuracy) was slightly outperformed by its oracle counterpart (81.53%). Table 6 reveals that a larger number of misclassifications led to deterioration than to improvement. However, the misclassified utterances resulting in improvements were on average longer and hence the number of tokens involved in improved and deteriorated performance, respectively, was approximately equal. The effect of misclassifications leading to improvements (+1.29%) was outweighed by those leading to deterioration (−1.52%), which ultimately resulted in the 0.23% absolute drop in performance.

For the multi-accent systems, Table 4 indicates that the parallel configuration (82.85%) yielded a slight improvement over its oracle counterpart (82.78%). Table 7 indicates that, although approximately the same number of misclassifications led to deteriorated and to improved performance, the number of tokens involved in improved performance was greater. As a result, the improvement as a result of misclassifications (+1.19%) was slightly larger than the deterioration (−1.12%), leading to the small overall improvement of 0.07%.

Table 8 presents the AID confusion matrix for the parallel multi-accent system employing per-utterance AID. The table indicates that confusions are most common between AE and CE, between AE and EE, and between CE and IE. Interestingly, such closeness between the CE and IE accents has recently also been established in an independent linguistic study.[29] The diagonal of Table 8 indicates that CE, EE and AE utterances are most prone to misclassification, while IE and BE are identified correctly more often. This analysis agrees with the pattern that was depicted in Figure 1, which highlights the similarity of AE, CE and EE, and the more distinct nature of BE and IE. The AID confusion matrix for the parallel per-utterance AID system using accent-specific models indicates similar trends.

Summary and conclusions

We investigated the effect of accent misclassifications on recognition accuracy when performing parallel speech recognition of the five accents of SAE. In order to isolate the effect of AID errors, the speech recognition performance of systems employing recognisers in parallel was compared with the performance of an oracle configuration in which each test utterance is presented only to the recogniser of the matching accent. Parallel configurations were also compared with accent-independent recognition achieved by pooling training data.

Our experimental results show that a parallel configuration applying AID at a per-utterance level and employing multi-accent acoustic models, which allow selective data sharing across accents, exhibited no degradation in accuracy compared to an oracle configuration despite a considerable number of AID errors. When AID was performed at a per-speaker instead of at a per-utterance level, we found that AID accuracy improved but that the recognition accuracy remained unchanged. An analysis of accent misclassifications indicated that misclassified utterances leading to improved speech recognition accuracy were on average longer than those leading to deteriorated accuracy. However, it was found that the majority (approximately 75%) of misclassified utterances did not affect speech recognition performance.

Table 8: Confusion matrix for the parallel per-utterance accent identification system using multi-accent acoustic models. Confusions are indicated as percentages (%).

		Hypothesised accent				
		AE	**BE**	**CE**	**EE**	**IE**
Actual accent	**AE**	62.41	3.48	17.42	11.32	5.37
	BE	4.16	78.26	7.38	3.36	6.85
	CE	16.50	6.77	53.88	8.60	14.25
	EE	21.94	3.70	7.27	58.83	8.26
	IE	2.43	7.40	10.98	7.75	71.45

AE, Afrikaans English; BE, Black South African English; CE, Cape Flats English; EE, White South African English; IE, Indian South African English.

We conclude that accent misclassifications occurring in a parallel recognition configuration do not necessarily impair speech recognition performance and that multi-accent acoustic models are particularly effective in this regard. This conclusion is important from the perspective of practical system implementation because it suggests that there is little to be gained from the inclusion of a more elaborate AID scheme prior to speech recognition. The inclusion of such an explicit AID component would significantly increase the design and implementation complexity of the overall speech recognition system. This increased cost would be particularly keenly felt in the under-resourced South African setting, in which suitable data and associated speech resources are scarce.

Acknowledgements

This work was executed using the High Performance Computer facility at Stellenbosch University. The financial assistance of the National

Research Foundation (NRF) of South Africa is acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

Authors' contributions

Both authors designed the experiments. H.K. performed the experiments and T.R.N. was the project leader. Both authors wrote the manuscript.

References

1. Statistics South Africa. Census 2011 report 03-01-41 [document on the Internet]. c2012 [cited 2013 Sep 06]. Available from: http://www.statssa.gov.za/census2011/default.asp
2. Odyssey 2010: The Speaker and Language Recognition Workshop; 2010 June 28 – July 01; Brno, Czech Republic. Baixas, France: International Speech Communication Association; 2010.
3. Schneider EW, Burridge K, Kortmann B, Mesthrie R, Upton C, editors. A handbook of varieties of English. Berlin: Mouton de Gruyter; 2004.
4. Finn P. Cape Flats English: Phonology. In: Schneider EW, Burridge K, Kortmann B, Mesthrie R, Upton C, editors. A handbook of varieties of English. Vol.1. Berlin: Mouton de Gruyter; 2004. p. 964–984.
5. Fukunaga K. Introduction to statistical pattern recognition. 2nd ed. San Diego, CA: Academic Press; 1990.
6. Kamper H, Muamba Mukanya FJ, Niesler TR. Multi-accent acoustic modelling of South African English. Speech Commun. 2012;54(6):801–813. http://dx.doi.org/10.1016/j.specom.2012.01.008
7. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings IEEE International Conference on Neural Networks; 1995 Nov 27 – Dec 01; Perth, Australia. IEEE; 1995. p. 1942–1948. http://dx.doi.org/10.1109/ICNN.1995.488968
8. Teixeira C, Trancoso I, Serralheiro A. Recognition of non-native accents. In: Proceedings of Eurospeech; 1997 Sep 22–25; Rhodes, Greece. Baixas, France: European Speech Communication Association; 1997. p. 2375–2378.
9. Fischer V, Gao Y, Janke E. Speaker-independent upfront dialect adaptation in a large vocabulary continuous speech recognizer. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP); 1998 Nov 30 – Dec 04; Sydney, Australia. p. 787–790.
10. Kirchhoff K, Vergyri D. Cross-dialectal data sharing for acoustic modeling in Arabic speech recognition. Speech Commun. 2005;46(1):37–51. http://dx.doi.org/10.1016/j.specom.2005.01.004
11. Despres J, Fousek P, Gauvain JL, Gay S, Josse Y, Lamel L, et al. Modeling northern and southern varieties of Dutch for STT. In: Proceedings of Interspeech; 2009 Sep 6–10; Brighton, UK. Baixas, France: International Speech Communication Association; 2009. p. 96–99.
12. Caballero M, Moreno A, Nogueiras A. Multidialectal Spanish acoustic modeling for speech recognition. Speech Commun. 2009;51:217–229. http://dx.doi.org/10.1016/j.specom.2008.08.003
13. Schultz T, Waibel A. Language-independent and language-adaptive acoustic modeling for speech recognition. Speech Commun. 2001;35:31–51. http://dx.doi.org/10.1016/S0167-6393(00)00094-7
14. Faria A. Accent classification for speech recognition. In: Renals S, Bengio S, editors. Proceedings of the Second International Workshop on Machine Learning for Multimodal Interaction (MLMI); 2005 July 11–13; Edinburgh, UK. Edinburgh: Springer; 2006. p. 285–293.
15. Chengalvarayan R. Accent-independent universal HMM-based speech recognizer for American, Australian and British English. In: Proceedings of Eurospeech; 2001 Sep 3–7; Aalborg, Denmark. Baixas, France: International Speech Communication Association; 2001. p. 2733–2736.
16. Teixeira C, Trancoso I, Serralheiro A. Accent identification. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP); 1996 Oct 3–6; Philadelphia, PA, USA. Philadelphia, PA: University of Delaware; 1996. p. 1784–1787.
17. Beattie V, Edmondson S, Miller D, Patel Y, Talvola G. An integrated multi-dialect speech recognition system with optional speaker adaptation. In: Proceedings of Eurospeech; 1995 Sep 18–21; Madrid, Spain. Baixas, France: European Speech Communication Association; 1995. p. 1123–1126.
18. Roux JC, Louw PH, Niesler TR. The African Speech Technology project: An assessment. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC); 2004 May 26–28; Lisbon, Portugal. Paris: European Language Resources Association; 2004. p. 93–96.
19. Stolcke A. SRILM – An extensible language modeling toolkit. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP); 2002 Sep 16–20; Denver, CO, USA. Denver, CO: Causal Productions; 2002. p. 901–904.
20. Katz SM. Estimation of probabilities from sparse data for the language model component of a speech recognizer. IEEE T Acoust Speech. 1987;35(3):400–401. http://dx.doi.org/10.1109/TASSP.1987.1165125
21. Ney H, Essen U, Kneser R. On structuring probabilistic dependencies in stochastic language modelling. Comput Speech Lang. 1994;8(1):1–38. http://dx.doi.org/10.1006/csla.1994.1001
22. Young SJ, Evermann G, Gales MJF, Hain T, Kershaw D, Liu X, et al. The HTK book (for HTK Version 3.4). Cambridge: Cambridge University Engineering Department; 2009.
23. Niesler TR. Language-dependent state clustering for multilingual acoustic modelling. Speech Commun. 2007;49(6):453–463. http://dx.doi.org/10.1016/j.specom.2007.04.001
24. Kamper H, Niesler TR. Multi-accent speech recognition of Afrikaans, Black and White varieties of South African English. In: Proceedings of Interspeech; 2011 Aug 28–31; Florence, Italy. Baixas, France: International Speech Communication Association; 2011. p. 3189–3192.
25. Bisani M, Ney H. Bootstrap estimates for confidence intervals in ASR performance evaluation. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2004 May 17–21; Montreal, Quebec, Canada. Montreal: IEEE; 2004. p. 409–412.
26. Mesthrie R. Socio-phonetics and social change: Deracialisation of the GOOSE vowel in South African English. J Socioling. 2010;14(1):3–33. http://dx.doi.org/10.1111/j.1467-9841.2009.00433.x
27. Bekker I, Eley G. An acoustic analysis of White South African English (WSAfE) monophthongs. South Afr Linguist Appl Lang Stud. 2007;25(1):107–114. http://dx.doi.org/10.2989/16073610709486449
28. Bekker I. Fronted /s/ in general White South African English. Lang Matters. 2007;38(1):46–74. http://dx.doi.org/10.1080/10228190701640025
29. Mesthrie R. Ethnicity, substrate and place: The dynamics of Coloured and Indian English in five South African cities in relation to the variable (t). Lang Var Change. 2012;24:371–395. http://dx.doi.org/10.1017/S0954394512000178