Automatic identification of phonetic similarity based on underspecification

Mark Kane[1], Julie Mauclair[2] and Julie Carson-Berndsen[1]

[1] School of Computer Science and Informatics, University College Dublin, Ireland
mark.kane@ucdconnect.ie, Julie.Berndsen@ucd.ie
[2] LIPADE, University Paris Descartes, France
julie.mauclair@parisdescartes.fr

Abstract. This paper presents a novel approach to the identification of phonetic similarity using properties observed during the speech recognition process. Experiments are presented whereby specific phones are removed during the training phase of a statistical speech recognition system so that the behaviour of the system can be analysed to see which alternative phone is selected. The domain of the analysis is restricted to specific contexts and the alternatively recognised (or *substituted*) phones are analysed with respect to a number of factors namely, the common phonetic properties, the phonetic neighbourhood and the frequency of occurrence with respect to a particular corpus. The results indicate that a measure of phonetic similarity based on alternatively recognised observed properties can be predicted based on a combination of these factors and as such can serve as an important additional source of information for the purposes of modelling pronunciation variation.
Keywords: speech recognition, phonetic similarity

.

1 Introduction

A key challenge in speech recognition is to construct acoustic models which correctly estimate a sub-word unit or phonetic class label within a specific time interval. The smallest posited linguistically distinctive unit that is typically modelled is the phoneme. However, phonemes that belong to the same acoustic-articulatory group (i.e. have similar acoustic or articulatory properties) are easily confused and thus statistical context-dependent phone models are often used as

[1] This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at University College Dublin and Dublin City University. The opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Science Foundation Ireland.

[2] The research presented in this paper is a revised and extended version of the paper presented at LTC 2009, Poznań Poland.

a basis for deciding which phoneme seems to be the best according to an acoustic probability and the probability of its occurrence with respect to a language model. Confusability of phonemes and the relationship to underlying phonetic properties of speech sounds remains an important area of research in order to address variability in the domain of speech recognition.

The research presented in this paper is motivated by research into patterns which emerge from mis-recognition of phonemes during the speech recognition process. The next step is to employ these mis-recognitions as a useful and additional source of information for modelling pronunciation variation.

Experiments presented in [1], demonstrated that almost 80% of all mis-recognised phonemes belong to the same phonetic group as the correct phoneme. In [2] this information is used in order to build *Broad Phonetic Groups* (BPG) that are defined according to a confusability matrix and a new phoneme classifier is proposed consisting of modular arrangements of experts, with one expert assigned to each BPG and focused on discriminating between phonemes within that BPG. The result in PER achieved by that system on the TIMIT corpus [3] is 26.4%. More recently these phonetic and phonological features are used to exploit similarities in phoneme recognition [4] with encouraging results. Neural Networks have also been used to classify between phonetic groups [5] and [6]. In [5] Neural Networks are used to improve the Discrete Wavelet Transform-based phonetic classification algorithm.

This paper focuses on the identification of phonetic similarity using properties observed during the speech recognition process. This approach also uses the notion of mis-recognitions but in a very different way from that outlined above. Rather than construct a confusability matrix for the recognised output, the output of two statistical speech recognition systems are compared, one where all phonemes to be recognised are included in the training data and one where individual phonemes are systematically removed from the training data (note that all phones occur in the testing data). This allows the identification of the *substituted* choice where a particular phoneme is not available as the number of substitutions will increase dramatically in the testing phase for the removed phoneme as there is no ASR (hidden Markov) model built to represent that phone from the training data. The domain of the analysis is restricted to specific contexts and thus the term *phone* will be used in the remainder of the paper. The substituted phones thus identified are then analysed with respect to their phonetic properties as given by a phonetic feature classification based on the IPA chart [7]. These properties provide a principled way to investigate phonetic similarity, underpinned by insights from experimental phonetics and phonological theory. The method described in this paper presents experiments on American English read speech and German spontaneous speech.

Section 2 presents the speech recognition system used in the experiment and section 3 details the experimentation carried out to identify the substituted phones. The results and analysis of the experiment are discussed in section 4 and 5. Some conclusions are drawn and directions for future work are highlighted in section 6.

2 Speech Recognition System and Corpus

The HMM based speech recognition system used in this experiment is implemented with HTK [8]. The TIMIT [3] and Kiel [9] speech corpora are used for training and testing of the HMM models. For completeness, the more technical details of the speech recognition system and the corpus are presented in the next section in the traditional way.

The TIMIT corpus consists of read speech spoken by 630 speakers of American English. The data is split into two sets; training and complete test set. The training set consists of 3696 utterances while the test consists of 1344 utterances. The SA data is not used in this paper. The Kiel corpus consists of spontaneous speech spoken by 26 speakers of German. The data is split into two sets; training and test set. The training set consists of 383 utterances while the test set consists of 142 utterances. There is no overlap between any of the training and test sets used in this paper. All plosives phones represented by a separate closure and associated burst are merged into a single phone.

The chosen form of parameterisation of a phone within an utterance is mel frequency cepstral coefficients (MFCCs), with their associated log energy and first and second order regression coefficients. Therefore every frame is represented by 39 coefficients. First, each speech waveform is passed through a pre-emphasis filter. The waveform is then framed at a rate of 10 ms with a frame size of 25 ms where each frame is then windowed using a Hamming window function.

These MFCCs representing the phones are then used in the calculation of the HMM models. The HMMs are context-dependent triphone models that were initially calculated by cloning and re-estimating context-independent monophone models. The triphones states were tied using phonetic feature decision trees for clustering. Each model is comprised of 5 states where only the centre 3 states are emitting. The decoding process is implemented with a uni-gram language model to emphasis acoustic variability thus negating the effects of a typical bi/tri-gram. Initially experimentation began with the evaluation of the decoding process using a trigram language model of the phones. However it was found that the substituting choice remained closer to the removed phones phonetic neighbourhood if a unigram language model was used. This process is executed without any pruning. Finally, the number of components in each mixture is set to 4 as this was found to be the optimal number for the Kiel corpus.

The experiment detailed in the next section distinguishes between a baseline and an underspecified system.

3 Experiment

In this paper, experiments based on an underspecified system for specific phones are presented. These experiments involve the identification of the substituted alternatives when one phone is removed from the training set. The phonetic make-up of each of the phones were classified with respect to features from the IPA chart.

As mentioned in the previous section, two types of speech recognition systems are evaluated: baseline and underspecified as shown in Figure 1. The baseline system is trained on a specific corpus training set. The underspecified system is trained on the same corpus training set but with all instances of a particular phone removed. Both systems are evaluated on the same test set.

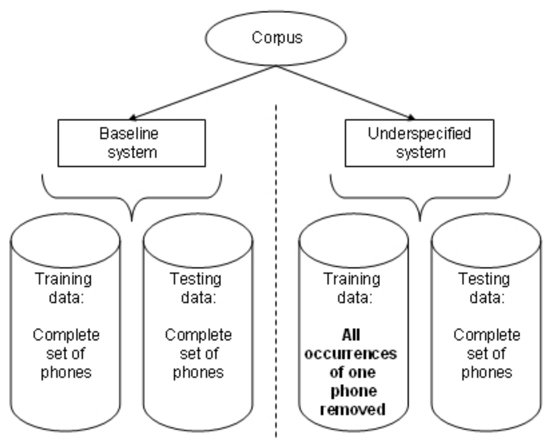

Fig. 1. Overview of baseline and underspecified systems.

From the baseline system all instances of the correctly recognised phones and all substitutions for the phones are identified. A certain level of control is required in these experiments and thus substitutions in which the left and right contexts of the substituted phone are correctly recognised are considered. This ensured that preceding errors had a limited affect on results. It is clear that for an underspecified system for a particular phone, there can be no correctly recognised instances of that phone. The substitution phones act as possible alternatives and serve as a comparator.

In this underspecified system where a phone is removed, it is anticipated that the substituted alternatives should be determined with respect to common phonetic properties, phonetic neighbourhood and perhaps with respect to frequency of occurrence in the full corpus, as this is an indicator of the relative frequency of the sound in the language (in particular the domain of the corpus). Each of these factors are underpinned by experimental phonetics and phonological theory (feature theory, markedness and underspecification, in particular [10]) and, in combination, should serve as indicators for phonetic similarity. The method is presented using the American English read speech corpus (TIMIT) and then the commonalities are outlined with the German spontaneous speech corpus (Kiel). Results have been calculated for all of the phones, but first the analysis of the phones [f] and [th] from the American English corpus are presented in sections 4 and 5.

4 Results

Based on the experiments outlined above, the results are presented as follows. Firstly, the results for the recognition of the phones [f] and [th] with the baseline system are presented. Secondly the results for the underspecified system for the phones [f] and [th] are presented and analysed in the context of phonetic similarity. Finally, the common phones between the American English and German corpora and their substitutions from an underspecified system are identified and discussed.

4.1 General information

The baseline system serves to highlight which substitutions are found for each of the phones when they are not correctly recognised, given a correct left and right context (LR context). Table 1 shows the general ASR statistics for these phones while Table 2 shows the top five substituted alternatives of these phones. The LR context quantity (a subset of the substitution quantity) is the number of the phones that were recognised incorrectly and were replaced by another phone, where both the phone to the left and right of the substituted phone were correctly recognised. In this table the following are also highlighted; the number of times a phone occurred, how many times it was recognised correctly, how many times it was not recognised correctly and was substituted, how many times it was not recognised but deleted altogether and how many times it was inserted. The sum of the recognised, substituted and deleted is equal to the occurrence of that phone within the test set. Tables 1 and 2 also contain the corresponding information for the underspecified system.

	Baseline		**Underspecified**	
Qty	f	th	f	th
occurrence	911	259	911	259
recognised	804 (88.3%)	88 (34%)	0 (0%)	0 (0%)
substituted	**94 (10.3%)**	**155 (59.8%)**	**708 (77.7%)**	**219 (84.6%)**
deleted	13 (1.4%)	16 (6.2%)	203 (22.3%)	40 (15.4%)
inserted	58	10	0	0
LR context	**23**	**42**	**206**	**60**

Table 1. General information of the baseline and underspecified (removed) system when evaluated against the test set.

As to be expected, it can be seen from Table 1 that the number of substitutions and LR contexts of a phone is greater in the underspecified system. This

	Substitutions (LR context)	
Reference	Baseline	Underspecified
f	th(6), p(4), t(3), s(2), d(2)	th(63), v(24), p(21), t(21), s(19)
th	f(17), t(9), dh(6), s(5), v(1)	f(18), t(11), dh(9), z(5), s(5)

Table 2. Top 5 substituted alternatives and quantities from the baseline and underspecified (removed) system with respect to LR context.

gives a broader range of similar phones as they are generated from a larger portion of data. This extra data allows a more complete pattern of substitutions to be observed.

5 Analysis of underspecified system

In this section the underlying theories are first presented with respect to particular examples. Therefore in order to illustrate the type of analyses undertaken, an underspecified system [f] and an underspecified system [th] from the American English corpus are analysed. This method of analysis was also performed on the full American English and German Corpora.

5.1 Phone [f]

As can be seen from Table 1, the baseline system, [f] was substituted 94 times and [th] was substituted 155 times. The number of [f] substitutions that have a correct LR context is 23 and the number of [th] substitutions that have a correct LR context is 42.

Due to the fact that in the underspecified system [f] and [th] are not correctly recognised, the amount of substitutions increased. In this system [f] was substituted 708 times and [th] was substituted 219 times. The number of [f] substitutions that have a correct LR context is 206 and the number of [th] substitutions that have a correct LR context is 60. This extra data helps disambiguate the substitution data as there is more of it to help ascertain a pattern. In this section, the following three factors are considered: 1. *common phonetic properties*, 2. *phonetic neighbourhood* and 3. *frequency of occurrence*.

5.2 Common phonetic properties of [f] and [th]

Phones can also be characterised in terms of the canonical properties or features they possess. Indeed this is typically one of the metrics employed to measure phonetic similarity. While the properties clearly relate to the notion of phonetic neighbourhood, depending on the feature set used, they may offer additional granularity which allows features to be grouped as natural classes. Table 3 provides the subset of relevant features for the phones of Table 2, where + means present and - means not present. Note that other subset groupings of these sounds may also be relevant for further experimentation.

Phone	**fric**	**plosive**	**bi-labial**	**labio-dental**	**dental**	**alveolar**	**voiced**
p	-	+	+	-	-	-	-
f	+	-	-	+	-	-	-
v	+	-	-	+	-	-	+
th	+	-	-	-	+	-	-
dh	+	-	-	-	+	-	+
s	+	-	-	-	-	+	-
t	-	+	-	-	-	+	-
d	-	+	-	-	-	+	+

Table 3. Subset of features describing phonetic properties of substituted phones for [f] and [th].

5.3 Phonetic neighbourhood

The notion of phonetic neighbourhood can be visualised as in figure 2 where the four planes of the cube indicated in the figure represent the fricative manner of articulation, the plosive manner of articulation, the voiced and the unvoiced articulations, where the horizontal dimension indicates the place of articulation within these planes. For the initial experiments, relative positions of phones on this cube were used as a basis for the analysis. A more detailed explication of the role of phonetic neighbourhood may be possible, based on a numerical distance measure and a differentiation between the relative weightings of the dimensions within the cube. For example, it may be that remaining on the *place* of articulation axis represents closer proximity than moving to the *voicing* axis for some occasions; it would certainly appear reasonable that a change in the manner of articulation should involve a greater phonetic distance in comparison to a change in voicing information. This is a topic for future work, however.

In the following subsections, the phonetic neighbourhoods of the phones [f] and [th] are discussed in more detail in the context of the substituted phones in Table 1 and 2.

Phonetic Neighbourhood of [f] The position of phone [f] should be regarded as the starting point in the cube of figure 2. The consonants which appear as the substitutions for [f] in Table 2 can be seen within the cube and their relationship is represented with respect to the four planes indicated in the figure and are as follows: unvoiced fricatives([th], [s]), voiced fricatives([v]) and unvoiced plosives([p], [t]).

At first glance, the LR context substitutes appear to be in rank order from Table 2. Cautiously, due to the small amount of data, the ranking for these substitutes may be partitioned into two groups: first - [th]; second - [v], [p], [t] and [s]. The phone [th] is the obvious first choice for substitution as its place of articulation on the fricative plane is closet to that of phone [f]. They also share the same manner of articulation and are both unvoiced. According to this depiction, moving one step in any dimension from [f] for example, phones [th], [v]

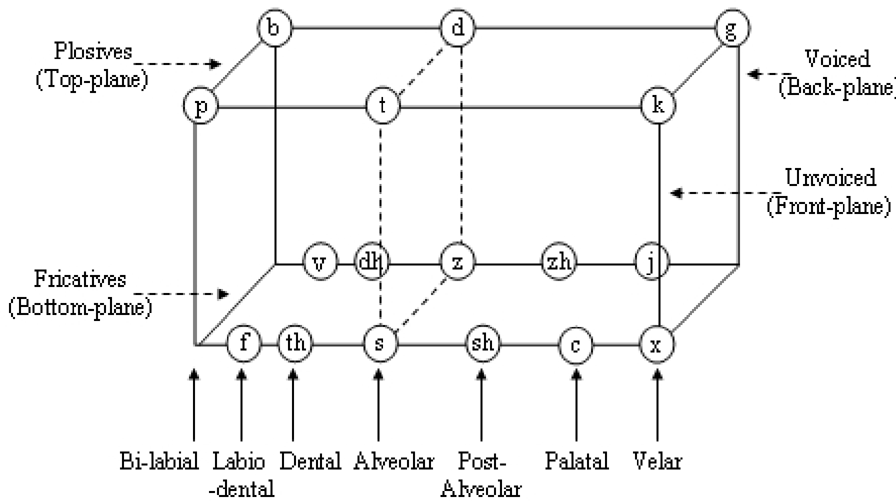

Fig. 2. Cube representation of phonetic neighbourhood.

and [p] would appear to have the closest phonetic neighbourhood to the phone [f]. The phonetic neighbourhood may be measured in terms of relative distance between points on the cube. Phonetic neighbourhood alone does not appear to be a complete determining factor (albeit close) for the substituted alternative for phone [f] as the phones [t] and [s] (where both phones are two places removed from the place of articulation of [f] and [t] has a different manner of articulation) seem to have as much priority as [v] and [p].

Phonetic Neighbourhood of [th] The notion of phonetic neighbourhood for [th] is also accounted for in figure 1. In this case, the position of phone [th] should be taken as the starting point in the cube.

Similar to the ranking consideration that was applied to the substitutions of phone [f], the ranking for the [th] substitutes from Table 2 may also be partitioned into two groups: first - [f]; second - [t], [dh], [z] and [s]. Again, the obvious first choice for substitution is [f] as it is also partially dental. Going one step in either dimension from [th] yields [f], [dh], [s] and almost [t]. Noticeably [t] seems to have as much priority as [dh] and may be attributed to the frequency of occurrence of [t] which will now be explained.

5.4 Frequency of occurrence

Another factor which is likely to be a determining factor for the substituted alternative is frequency of occurrence in the corpus as a whole. This factor relates to the notion of markedness in phonology which postulates that the most unmarked (or default) sound in a language is also likely to be the sound which is most common. The frequencies of occurrence of the phones in the full American English corpus (as described in section 2) together with their associated percentage of the corpus they represent are presented in Table 4.

Phone	Frequency of phone in full TIMIT corpus	Percentage of phone in full corpus
t	8578	5.16%
s	8348	5.03%
d	5918	3.56%
p	4015	2.42%
dh	3272	1.97%
f	3126	1.88%
v	2704	1.63%
th	1004	0.60%

Table 4. Frequencies of phones of Table 2 from the combined training and test set of the American English corpus.

These frequency of occurrences can also be visualised with respect to the phonetic neighbourhood cube as depicted in figure 3, where the larger the circle, the greater the frequency of occurrence. As mentioned in section 5.3, [t] and [s] seem to have as much priority as [v] and [p] as substitutions for [f] and may be attributed to their frequency of occurrence. A region of high frequency of occurrence within the cube has a pseudo gravitational effect, the closer to this region, the greater the influence this region has on the substituting possibilities; creating a possible default substitution combined with unmarked sounds [10].

The common phonetic properties, the phonetic neighbourhood and the frequency of occurrence in the corpus, together determine how the results of the underspecified systems for [f] and [th] should be interpreted. The corresponding substitutions of phone [f] for German using the same method are noted in Table 5.

5.5 Discussion

In this section the substituted phones for the underspecified systems of both the American English and German speech corpora are presented, where the complete set of results for the phones that are common to both corpora are listed in Table 5 with their associated substitutions. However it is important to note that some of the substituting phones are exclusive to one corpus. Also the substituted phones are displayed left to right with a decreasing substitution count.

One of the first items of note is that the [ɨ] phone of the American English corpus (known as [ix]) is a frequent substituting phone. The reason for this is clarified when its frequency of occurrence is found to be the highest for its

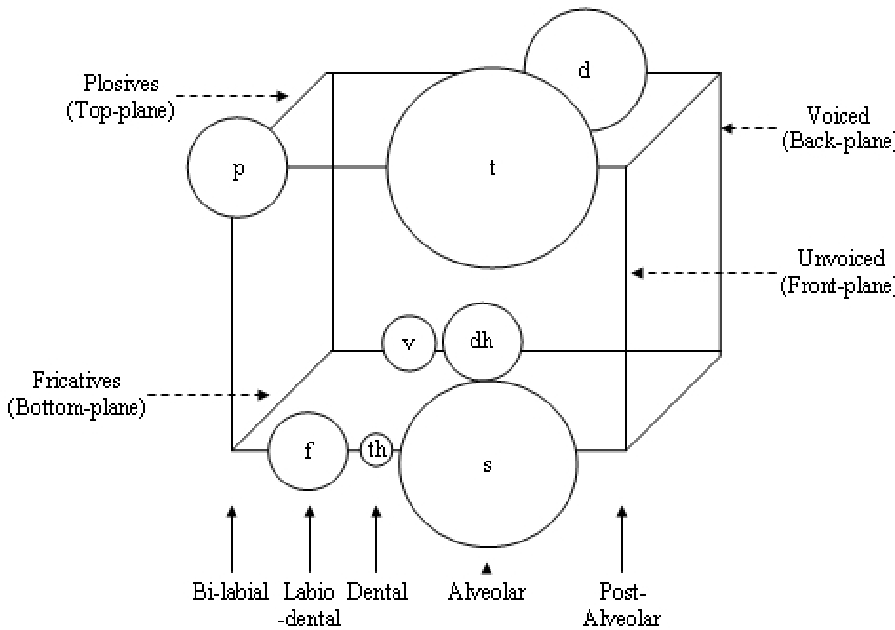

Fig. 3. Representation of phonetic neighbourhood with respect to the most frequent phones from the American English corpus, where the larger the circle the greater the frequency of occurrence.

corpus, over twice that of the schwa ə. This close-central vowel appears to be the default sound for the domain of the American English corpus as it substitutes both vowel and consonants alike and appears as a top five substitute for several phones as well as appearing as a substitute beyond the top five list.

Looking at [f] from the German corpus it can be seen that in one step in either dimension within the phonetic neighbourhood cube of figure 2 the obvious voiced step [v] and place of articulation step [s] are found. Plosive phones [t] and [d] of the same place of articulation of [s] are also noted. It is important to note that the [θ] and [ð] phones known as [th] and [dh] are exclusive to the American English corpus. Therefore when understanding [f]'s substitutions for the German corpus, [th] and [dh] are not a substituting choice. Looking at these results with respect to their frequency of occurrence, it was found that while [t] substituted [f] the most, it also has the highest frequency of occurrence (percentage of phone in the complete German corpus) of (9.1%). [v] ranked as the second most substituted phone with a frequency of occurrence of 2.85% where [s] ranked the third most substituted phone even though its frequency of occurrence is 6.24%. In this case the phonetic neighbourhood similarity between [f] and [v] outweighed the frequency of occurrence of [s] but not [t]. Finally the unvoiced velar plosive [k] with a frequency of occurrence of 1.7% outweighed the voiced alveolar plosive [d] with its frequency of occurrence of 5.07%. It appears that the choice of substituting phone is based on primarily two weights, (phonetic neighbourhood proximity of substituting phone)*(frequency of occurrence of substituting phone).

	HMM models		substitutions			HMM models		substitutions	
IPA	**TIMIT**	**Kiel**	**TIMIT**	**Kiel**	**IPA**	**TIMIT**	**Kiel**	**TIMIT**	**Kiel**
p	p	p	t,b,k,f,d	t,v,b,g,k	ə	ax	@	ɨ,ʌ,ɪ,ɛ,oʊ	n,ɪ,t,d,ɛ
b	b	b	d,p,v,ð,ə	d,t,v,m,a	i	iy	i:	ɨ,eɪ,ɾ,ɪ,y	e,l,n,ɪ,t
t	t	t	d,k,p,s,tʃ	d,k,v,p,s	y	y	y:	i,ɨ,ʉ,ɪ,k	ə,v,ɹ,i,k
d	d	d	t,b,ð,g,ɨ	t,n,z,g,ə	ɪ	ih	I	ɨ,ɛ,i,eɪ,ə	ə,ɛ,e,i,l
k	k	k	t,g,p,d,ɨ	t,g,p,x,ɟ	u	uw	u:	ʉ,ə,oʊ,ɨ,l	ʊ,o,a
g	g	g	k,d,b,t,ɨ	d,k,z,t,b	ʊ	uh	U	ə,ɨ,ʌ,oʊ,ɛ	u,ɔ,a,ɤ,n
f	f	f	θ,v,p,t,s	t,v,s,k,d	ɛ	eh	E,E:	ɪ,æ,ʌ,ɨ,eɪ	ɪ,ə,a
v	v	v	f,b,θ,p,d	f,d,m,b,t	ɔ	ao	O	ɑ,oʊ,w,l,ʌ	a,aʊ,ə,n
s	s	s	z,t,ʃ,f,ɨ	t,f,z,ç,a	aʊ	aw	aU	æ,ɑ,ɔ,ʌ,oʊ	a,ɔ,ʊ,ə,o
z	z	z	s,dʒ,ʃ,ð,ɨ	s,t,d,n,l	aɪ	ay	aI	ɑ,ʌ,æ,eɪ,ɛ	a,ɐ,ə,l,ɛ
m	m	m	n,ɨ,ə,v,b	n,ə,l,z,ɪ	l	l	l	ə,w,oʊ,ɔ,ɹ	n,t,ɪ,d,m
n	n	n	m,ŋ,ɨ,d,t	ə,m,t,ɪ,ʊ	ŋ	ng	N	n,m,ɨ,g,k	n,k,u,o

Table 5. Subset of American English and German phone substitutions (LR context) with respect to the IPA using an underspecified system. Only the top five substitutions are shown.

6 Conclusion

This paper has presented a novel approach to the identification of phonetic similarity using properties observed during the speech recognition process. An experiment was presented whereby the phones [f] and [th] were separately removed during the training phase of a statistical speech recognition system so that the behaviour of the system could be analysed to see which alternative phones were selected. The analyses of underspecified systems was illustrated using the phones [f] and [th] from an American English corpus with respect to three determining factors: common phonetic properties, phonetic neighbourhood and frequency of occurrence of the phone in the full corpus. Other common phones to both the American English and German corpora yielded similar results and are presented in Table 5.

The frequency of occurrence of a substituted phone was seen to have a strong effect on its ranking among the alternatives although phonetic neighbourhood and phonetic properties also played an important role. In the context of the results and analysis presented in section 4 and 5, phonetic neighbourhood was interpreted without resort to an exact measure of distance between the points on the cube. As mentioned there, a more detailed explication of the role of phonetic neighbourhood may be possible, based on a numerical distance measure and a differentiation between the relative weightings of the dimensions of the cube. For example, it may be that remaining on the *place* of articulation axis represents closer proximity than moving to the *voicing* axis for some occasions; it would certainly appear reasonable that a change in the manner of articulation would involve a greater phonetic distance in comparison to a change in voicing information.

A factor which will be taken into account in future work is the extent to which the context (prosodic position and segmental context [10]) influences the substituted alternative; for example a substituted phone may have emerged as a result of an influence of the preceding or following context/phone.

All of these points will be considered in the next phase of experimentation and the experiments will be extended to include other broad classes of phones with the aim of providing a principled methodology for the prediction of phonetic similarity for the purposes of speech recognition. A phonetic similarity measure can serve as an important additional source of information for the construction of acoustic models in statistical speech recognition, for enhancing the lexicon with appropriate phonetic variants and for the design of knowledge-based feature detection engines. In summary, future work envisages that these phonetic similarity measures will be used with the output of an ASR system so that if a confidence metric is low for some output, then the phonetic similarity measure will enable the system to offer an appropriate alternative.

References

1. Halberstadt, A. and J. Glass, 1997. Heterogeneous acoustic measurements for phonetic classification. *Eurospeech proceedings*:pp 401–404.
2. Scanlon, P., D. Ellis, and R. Reilly, 2007. Using broad phonetic group experts for improved speech recognition. *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 3:pp 803–812.
3. Garofolo, J., L. Lamel, W. Fisher, J. Fiscus, D. Pallett, and N. Dahlgren, 1993. The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CDROM.
4. Mauclair, J., D. Aioanei, and J. Carson-Berndsen, 2009. Exploiting phonetic and phonological similarities as a first step for robust speech recognition. *EUSIPCO proceedings*.
5. Thuan, P. Van and G. Kubin, 2005. Dwt-based phonetic groups classification using neural networks. *ICASSP proceedings*:pp 401–404.
6. Ghiselli-Crippa, T. and A. El-Jaroudi, 1991. Voiced-unvoiced-silence classification of speech using neural nets. *IJCNN proceedings*:pp 851–856.
7. The-International-Phonetic-Alphabet, 2005. *http://www.langsci.ucl.ac.uk/ipa/*.
8. Young, Steve, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Xunying (Andrew) Liu, Gareth Moore andJulian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, and Phil Woodland, 2009. Hidden markov model toolkit (htk). *http://htk.eng.cam.ac.uk/*, Version 3.4.1.
9. IPDS, CD-ROM#2: The Kiel Corpus of Spontaneous Speech , Vol. 1, Kiel, IPDS, 1995.
10. Chomsky, N. and M. Halle, 1968. The sound pattern of english. *Harper & Row*.