interspeech 2008 incorporating SST 08

Distinctive Feature Fusion for Recognition of Australian English Consonants

Trent W. Lewis and David M.W. Powers

School of Computer Science, Engineering and Mathematics, Flinders University, Australia

[trent.lewis|david.powers]@flinders.edu.au

10.21437/Interspeech.2008-662

Abstract

Audio-Visual Automatic Speech Recognition offers to make speech recognition possible in noisy environments. Early and late fusion approaches dominate the field but may ignore linguistically relevant features. *Distinctive features* offer an alternative unit for fusion and research has shown that this is feasible on subsets of phonemes [1]. This paper outlines two extended models, multi-class and binary, and results suggest that it is possible to achieve a 20dB gain over audio-only recognition in low SNR environments.

Index Terms: audio-visual fusion, distinctive features

1. Introduction

The objective of Audio-Visual Automatic Speech Recognition (AV-ASR) is to enhance traditional speech recognition by incorporating a visual signal into the system [2]. A simple way to achieve this is to combine both the acoustic and visual features into one large feature vector which is used for recognition. This technique is effective, given enough training data, time and resources, but we can use knowledge from psychology and linguistics to conceive a more elegant and effective system.

Visually perceivable speech gestures group into distinct classes of phoneme-like visemes that are complementary to speech sounds difficult to perceive in high acoustic noise [3]. Sub-systems can thus be specialised for their modality and increase the overall system accuracy [1]. However, a one-to-many mapping does not exist between visemes and phonemes, so it may add another layer of complexity.

Early, or feature fusion is when the acoustic and visual features are concatenated together and then classified. Late, or decision fusion is when each modality is classified separately and the classification outputs are fused. The main unit for classification is the phoneme and late fusion is usually achieved using a weighted product of the outputs:

$$P(w_i) = \prod^{s\in\{A,V\}} P(w_i|s)^{\lambda_s} \quad (1)$$

where $P(w_i)$ is the estimated probability of phoneme w_i, s is the modality of either audio (A) or video (V), λ_s is the weight associated with modality s, and $P(w_i|s)$ is the conditional probability of w_i given modality s. By incorporating relative stream reliability (λ_s), late fusion has the potential to allow more accurate classification, increased generalisation and faster more reliable training, but this potential has not generally been realised and does not take into account any linguistic insights into the abilities of the different modalities.

This paper investigates the linguistic *distinctive feature* as an alternative unit of AV fusion. The approach of distinctive feature fusion is explained and then a set of experiments comparing performance with the standard architectures are presented.

2. Distinctive Feature Models

A distinctive feature is a phonetic *feature* that *distinguishes* minimal pairs of words for a particular language [4]. For example, the difference (ignoring aspiration) between the minimal pair /pit/ and /bit/ is that the initial sound differs in the feature voice ($-$voice for /pit/ and $+$voice for /bit/). Thus, we can say that the phonemes /p/ and /b/ are distinctive sounds in English and also that the feature [$\pm$voice] is a *distinctive feature*. Phonemes are usually considered the base element in most speech recognition systems, however, they are actually a mask for a bundle of distinctive features [4]. Thus, the minimal units for speech perception are arguably the distinctive features for a language.

In this research we investigate distinctive feature fusion and compare its performance to early and late fusion models. Previous research has highlighted distinctive features as a promising approach [5, 6, 7]. Distinctive feature fusion was presented in [1] and showed advantages for nine English plosives, /p,b,m,t,d,n,k,g,ng/, which vary on two specific distinctive features that are more easily distinguished acoustically (voicing: unvoiced/voiced/nasal) or visually (place: labial/-alveolar/velar).

This model of distinctive feature fusion is outlined in Figure 1. In comparison to early and late fusion, this model has the classification unit of a particular distinctive feature (DF), which in this case is either voicing or place. This model can be expressed in three different forms: *1. DF-I*: audio influences voicing and video the place, *2. DF-II*: audio can influence both voicing and place, *3. DF-III*: audio and video influence both features. DF-I is similar to the VPAM model [5], whilst the other two allow greater influence of each modality. In our current experiments competitive fusion (a crossed circle) uses the product (or geometric mean) of the modalities, whilst complementary fusion (an uncrossed circle) integrates the fused outputs of the place and voice classifiers and uses the sum (or arithmetic mean).

The simplistic three class classifications of place and voice were satisfactory for the nine phoneme class distinctive feature classification, but extending the concept of distinctive feature fusion to 22 English consonants requires a reconfiguration of the base classifiers as it is not possible to identify each consonant individually using the two classifiers.

3. Extended Distinctive Feature Models

Two options are explored for the implementation of the distinctive feature fusion architecture. The first extends the current architecture by including more classes into the place classifier and introducing a manner of articulation (manner) classifier. The manner of articulation reflects the way the air stream is affected as it travels from the lungs up and out of the mouth and nose. Thus, a phoneme is now determined from the complementary

Accepted after peer review of full paper

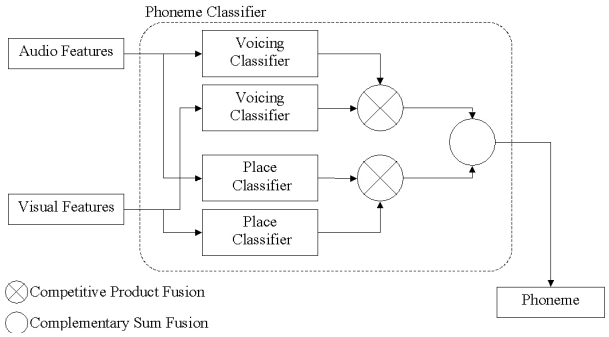

Figure 1: Distinctive Feature Fusion Model

Table 1: The class labels for the PVM and Binary feature classifiers. p = place, v = voicing, m = manner, cs = consonantal, s = sonorant, n = nasal, v = voiced, ct = ct, l = labial, a = alveolar, p = palatal, vl = velar, and sb = sibilant.

	PVM			Binary								
	p	v	m	cs	s	n	v	ct	l	a	p	vl	sb
p	1	1	1	1	0	0	0	0	1	0	0	0	0
b	1	2	1	1	0	0	1	0	1	0	0	0	0
t	2	1	1	1	0	0	0	0	0	1	0	0	0
d	2	2	1	1	0	0	1	0	0	1	0	0	0
k	6	1	1	1	0	0	0	0	0	0	0	1	0
g	6	2	1	1	0	0	1	0	0	0	0	1	0
f	3	1	2	1	0	0	0	1	1	0	0	0	0
v	3	2	2	1	0	0	1	1	1	0	0	0	0
th	4	1	2	1	0	0	0	1	0	0	0	0	0
dh	4	2	2	1	0	0	1	1	0	0	0	0	0
s	2	1	2	1	0	0	0	1	0	1	0	0	1
z	2	2	2	1	0	0	1	1	0	1	0	0	1
sh	5	1	2	1	0	0	0	1	0	0	1	0	1
ts	2	1	3	1	0	0	0	0	0	1	0	0	1
dz	2	2	3	1	0	0	1	0	0	1	0	0	1
m	1	3	1	1	0	1	1	0	1	0	0	0	0
n	2	3	1	1	0	1	1	0	0	1	0	0	0
ng	6	3	1	1	0	1	1	0	0	0	0	1	0
l	2	2	4	1	1	0	1	1	0	1	0	0	0
r	2	2	5	0	1	0	1	1	0	1	0	0	0
w	6	2	5	0	1	0	1	1	0	0	0	1	0
j	5	2	5	0	1	0	1	1	0	0	1	0	0

fusion of three base classifiers place, voice, and manner, the classes of which are; *place*: labial (1), alveolar (2), labio-dental (3), inter-dental (4), palatal (5), velar (6), *voice*: unvoiced (1), voiced (2), nasal (3), *manner*: stop (1), fricative (2), affricate (3), lateral (4), glide (5). For example, the phoneme /s/ is an alveolar, unvoiced, fricative and differs from /t/ only in its manner (/t/ is a stop). Table 1 presents each of the phonemes and their distinctive feature classifications used in the following experiments. This configuration is referred to as PVM.

The second approach taken for distinctive feature classification and fusion is to assess each phoneme on a set of *binary* distinctive features. The first approach groups particular distinctive features into a set of natural multi-class classifications. However, the original formulation of distinctive feature theory is that a phoneme is simply an alternative representation to the underlying bundle of binary features, and it is these features that enable the listener to distinguish different linguistic sounds [4]. Therefore, a binary feature representation may facilitate a better level for the fusion of underlying audio-only and video-only classifiers and thus lead to an increased overall phoneme recognition performance.

The formulation taken here allows the distinction between the 22 phonemes and consists of the following ten binary features: *Consonantal*, *Sonorant*, *Nasal*, *Voiced*, *Continuant*, *Labial*, *Alveolar*, *Palatal*, *Velar*, and *Sibilant*. How these features relate to each of the 22 phonemes is outlined in Table 1. The output of ten binary classifiers are mapped to a particular phoneme by concatenating the outputs and finding the distinctive feature set that is the closest to the concatenated feature set. In the tables and figures that follow, this binary distinctive feature arrangement is labelled as Binary.

To achieve the three distinctive feature fusion models the features that have been found to be robust to acoustic noise in human perception and more easily perceivable visually are classified by the audio-only and video-only classifiers, respectively. For example, Benki [8] recently confirmed that voicing and manner characteristics are relatively robust to noise, whilst place of articulation is severely affected. Thus, for the PVM configuration the video-only classifier is responsible for only the place classification in the DF-I and DF-II models. The audio-only classifier is then responsible for both the manner and voicing classifications in the DF-I model (in the style of VPAM [5]) and in the DF-II model it can also influence the place classification. Similarly, in the Binary case the video-only classifiers make judgments for the Labial, Alveolar, Palatal, and Velar distinctive features which are all place of articulation based features. The remaining distinctive features are mainly characterised by manner and voicing distinctions and are thus in the domain of the audio-only classifier. In the DF-III model both the audio-only and video-only classifiers for all the features contribute to the final classification.

4. Experimental Evaluation

Base Classifiers: The base classifiers (audio-only, video-only, early fusion, 3 multi-class, and 10 binary) are three-layer ANNs trained using back propagation with an adaptive learning rate and momentum. Training is stopped early if there is large deviation in performance on the validation data and a ten fold cross-validation method is used. For more information on classifier parameters and results see [9]. The AVOZES corpus, module 4, was used to evaluate the proposed models where there are 20 examples of each of the 22 consonants from 20 native Australian Speakers [10]. Visual features were mouth height and width and relative teeth count and first 12 Mel-cepstrum coefficients, the log-power, their deltas as acoustic features. Performance was assessed using the bookmaker measure, $[0 \ldots 1]$, which gives the probability of randmonly guessing (0) as opposed to making an informed decision (1) [11].

The overall performance of the place and voicing classification was as expected for the audio- and video-only based classifiers. The audio-only voicing classification is superior to the video-only voicing and when compared to the other audio-only PVM classifications. On the other hand, the video-only place classifier is superior to the overall audio-only place classifier. The audio-only manner classifier is better than the video-only, but it is inferior to the audio-only voicing classifier, whilst in the video-only context place is better than manner and manner better than voicing classification. The best classified binary features, in decreasing order, are voiced, palatal, sibilant, and sonorant. The voiced feature again showed its resilience to additive noise, displaying a similar shape to the voicing classifier of the PVM configuration. The palatal binary feature, which here has been grouped as a place feature and thus more easily identified visually, is the second best classified audio-only feature. How-

ever, this may be because this class only contains two positive classes of /sh/ and /j/ and thus it is conceivable that other factors are influencing the classification of this binary feature. Fortunately, the labial feature is better classified visually, surpassing the audio-only classification even in clean acoustic conditions, and the video-only alveolar binary feature is also competitive with the audio-only in the overall performance values.

Audio-Visual Fusion: Table 2 presents the results of the audio-only and fusion phoneme classification performance for the 22 consonants. A one-way ANOVA for each of the noise levels 1dB, 15dB, 25db, 35dB and clean revealed that there are significant differences between the fusion models, F(17, 167) = 213.34, 180.32, 201.31, 216.25, 186.93, $p < 0.05$, respectively. Multi-way post-hoc comparisons revealed the specific differences between the means of each of the fusion models. In Table 2 an α indicates that the mean is significantly different from audio-only, a β indicates a significant difference from unweighted late fusion, and a γ indicates a significant difference from the optimally weighted late fusion.

The early fusion is consistently better than the audio-only, but this difference is only significant on the clean data ($p < 0.05$). This results in a overall relative error reduction (column labelled "All") of 4.38% from the audio-only. Examining the results of the unweighted late fusion, however, reveals that this common fusion method has a negative, albeit insignificant, relative error reduction. Indeed, for the 35dB and clean data the performance of the late fusion is significantly worse than the audio-only classifier, thus exhibiting catastrophic fusion. Moreover, given that the video-only phoneme performance is 0.105, both the early and late fusion models are catastrophically fusing at the 1dB level.

Unweighted: The unweighted PVM, DF-I model, on the other hand, is always significantly better than the late fusion model and significantly better than the audio-only model for lower SNRs of 1dB, 5dB, and 25dB. This results in relative error reduction of 9.60% which is the highest overall error reduction for all of the unweighted fusion models and configurations. Interestingly, the DF-II model was the best of the unweighted models for the plosive subset, but for the PVM configuration it has the lowest reduction. In this case, the addition of the audio-only place classifier is impeding the performance in the DF-II model. Indeed, with the addition of this classifier the DF model performance drops below that of the video-only phoneme classifier at the 1dB level. Conversely, allowing the video-only voicing and place classifiers to interact in the DF-III model improves performance from the DF-II model with no significant catastrophic fusion. As the performance of the video-only voicing classifier is close to zero, this advantage of the DF-III classifier must be due to the influence of the video-only manner classifier, especially in the lower SNRs. Overall, however, the error reductions for the PVM, DF-II and -III models are no greater than for the early fusion.

The Binary DF models, however, show a different pattern of results. The overall error reduction for the Binary DF-II is better than the DF-I which can be attributed to the poor performance of the video-only classifiers on the place based binary features and the superior performance of the corresponding audio-only classifier that are able to influence these binary features in the DF-II model. The Binary, DF-III model is only slightly better overall than the DF-I, but both have significant catastrophic fusion above 30dB when compared to the audio-only classifier and perform worse than the video-only at 1dB.

Neurally Weighted: In the previous work on confidence estimation it was shown that using an ANN to learn the fusion of the audio- and video-only classifier outputs was the best approach to fusion [9]. In light of this, the group labelled "fusion, neurally weighted" in Table 2 uses an ANN to fuse the two different modalities such that each modality was able to influence each of the features (as in the DF-III model). The ANN had 20 neurons on the hidden layer and was trained for 500 epochs using back propagation with an adaptive learning rate and momentum. The ANN was trained using the validation data with either only the clean data (labelled "Clean") or with the addition of the 5dB noisy data (labelled "Noisy"). The cleanly trained fusion ANN did not perform to even the same level as the unweighted fusion, especially for the PVM configuration. For example, at 35dB the PVM, Clean model is significantly worse than the audio-only and worse than the video-only at 1dB. The cleanly trained Binary fusion has a similar pattern of performance except that it is also catastrophically fusing on the clean test data. The noisy trained Binary fusion also has significant catastrophic fusion on the clean data, but has superior performance on lower SNRs and thus has the best Binary overall error reduction of 6.40%. Similarly, the noisy trained PVM fusion produces the best overall relative error reduction out of the unweighted and neurally weighted fusion models. This superior error reduction is due to its performance being significantly better than audio-only and unweighted late fusion performance in SNRs 25dB and less.

Optimally Weighted: The last group in Table 2 presents the results from optimally (in the minimum oracular sense) weighting the audio- and video-only classifiers in the late fusion, DF-II and -III models. The first point of interest here is that the neurally weighted PVM fusion using noisy data for training is not significantly worse than the optimal late fusion at a SNR of 1dB. Thus, it may be possible in practice to reach the limits of the optimal fusion. Both the optimal PVM and Binary variants of the DF-II model produce overall performance better than unweighted and neurally weighted models, but optimally weighting these DF-II variants does not exceed the performance of the optimally weighted late fusion and are indeed significantly less for most of the SNRs tested. The best optimal performance was found for DF-III architectures.

Note that, it is the Binary, DF-III configuration that gives the best overall error reduction of 48.82%. This is interesting as the Binary feature configuration has lagged behind the PVM in the other experimental results, and this validates the Binary feature configuration as a viable model for phoneme recognition. That is, if it is possible to estimate the *a posteriori* probabilities of each of the binary features (either audio-only or audio-visually) then their complementary combination can lead to a superior phoneme classification.

Low SNR: The main reason that the visual modality is incorporated into a speech recognition system is to assist in the adverse, low acoustic SNR environments. The final column in Table 2 gives an indication of the gain for only the low SNRs tested (< 10dB). Here, a slight change in the pattern of results can be seen, but, most importantly, the advantage of the DF fusion models over the early and late models is further emphasised. Indeed, the early fusion and the optimally weighted late fusion models actually decrease in their error reduction when only the low SNRs are considered. This indicates that the majority of the error reduction for these models is in the higher SNRs where the improvement is not necessarily required as the audio-only classifier is usually far superior, and the visual data can potentially only make minor improvements. In contrast, all the DF models have a larger error reduction than early and unweighted late fusion when only considering the lower SNRs.

Table 2: Performance of audio-only, unweighted and optimally weighted fusion methods. α, β, and γ indicate a significant difference from audio-only, late unweighted or late weighted, respectively ($p < 0.05$). Error Red. is the relative error reduction from audio-only: All is overall and <10dB only includes reductions from SNRs less than 10dB.

Fusion	SNR (dB)					Error Red. (%)	
Method	1	15	25	35	clean	All	<10dB
audio-only	0.054^{γ}	0.171^{γ}	0.284^{γ}	$0.392^{\beta\gamma}$	$0.443^{\beta\gamma}$	0.00	0.00
fusion, unweighted							
early fusion	0.078^{γ}	0.202^{γ}	0.315^{γ}	$0.420^{\beta\gamma}$	$0.484^{\alpha\beta\gamma}$	4.38	3.22
late fusion	0.086^{γ}	0.181^{γ}	0.281^{γ}	$0.348^{\alpha\gamma}$	$0.387^{\alpha\gamma}$	-0.87	3.23
PVM, DF-I	$0.182^{\alpha\beta\gamma}$	$0.291^{\alpha\beta\gamma}$	$0.331^{\alpha\beta\gamma}$	$0.391^{\beta\gamma}$	$0.447^{\beta\gamma}$	9.60	16.62
PVM, DF-II	0.090^{γ}	0.201^{γ}	0.295^{γ}	0.378^{γ}	$0.443^{\beta\gamma}$	2.42	5.21
PVM, DF-III	$0.111^{\alpha\gamma}$	0.213^{γ}	0.294^{γ}	0.380^{γ}	$0.429^{\beta\gamma}$	2.89	6.67
Binary, DF-I	$0.124^{\alpha\gamma}$	0.214^{γ}	0.271^{γ}	$0.327^{\alpha\gamma}$	$0.371^{\alpha\gamma}$	0.05	8.00
Binary, DF-II	0.091^{γ}	$0.235^{\alpha\beta\gamma}$	0.310^{γ}	$0.389^{\beta\gamma}$	$0.430^{\beta\gamma}$	4.21	5.97
Binary, DF-III	0.085^{γ}	$0.224^{\alpha\gamma}$	0.277^{γ}	$0.335^{\alpha\gamma}$	$0.374^{\alpha\gamma}$	0.41	5.65
fusion, neurally weighted							
PVM, Clean	$0.099^{\alpha\gamma}$	0.196^{γ}	0.269^{γ}	$0.341^{\alpha\gamma}$	0.413^{γ}	0.07	6.00
PVM, Noisy	$0.242^{\alpha\beta}$	$0.297^{\alpha\beta\gamma}$	$0.340^{\alpha\beta\gamma}$	0.378^{γ}	0.421^{γ}	9.83	20.36
Binary, Clean	0.087^{γ}	$0.224^{\alpha\gamma}$	0.298^{γ}	$0.345^{\alpha\gamma}$	$0.375^{\alpha\gamma}$	1.14	5.25
Binary, Noisy	$0.158^{\alpha\beta\gamma}$	$0.278^{\alpha\beta\gamma}$	$0.337^{\alpha\beta\gamma}$	0.364^{γ}	$0.373^{\alpha\gamma}$	6.40	13.38
fusion, optimally weighted							
late fusion	$0.265^{\alpha\beta}$	$0.390^{\alpha\beta}$	$0.497^{\alpha\beta}$	$0.572^{\alpha\beta}$	$0.616^{\alpha\beta}$	27.71	23.25
PVM, DF-II	$0.235^{\alpha\beta}$	$0.351^{\alpha\beta}$	$0.390^{\alpha\beta\gamma}$	$0.456^{\alpha\beta\gamma}$	$0.519^{\alpha\beta\gamma}$	17.94	22.51
PVM, DF-III	$0.482^{\alpha\beta\gamma}$	$0.551^{\alpha\beta\gamma}$	$0.597^{\alpha\beta\gamma}$	$0.645^{\alpha\beta\gamma}$	$0.667^{\alpha\beta\gamma}$	44.46	46.33
Binary, DF-II	$0.213^{\alpha\beta\gamma}$	$0.336^{\alpha\beta\gamma}$	$0.411^{\alpha\beta\gamma}$	$0.483^{\alpha\beta\gamma}$	$0.523^{\alpha\beta\gamma}$	18.38	19.29
Binary, DF-III	$0.485^{\alpha\beta\gamma}$	$0.593^{\alpha\beta\gamma}$	$0.633^{\alpha\beta\gamma}$	$0.677^{\alpha\beta\gamma}$	$0.700^{\alpha\beta\gamma}$	48.82	48.25

Indeed, the unweighted PVM, DF-I has an error reduction of 16.62% which is twice that of any of the other unweighted fusion models. Moreover, the noisy trained, neurally weighted PVM configuration has an error reduction of 20.36 which is actually competitive with several of the optimal fusions. This indicates that, unlike the early and late fusion models, the DF models are able to reduce the error in the SNR range where it matters most for boosting the recognition performance (i.e. the more noisy conditions).

5. Summary and Discussion

The purpose of this paper was to reevaluate the concepts and ideas from previous work on an independently collected and more challenging corpus. The AVOZES data corpus was appropriate for this purpose in that it was multi-speaker in nature and covered the most common phonemes used in Australian English. Extending the distinctive feature model to the full consonant set proved that this model is a worthwhile pursuit especially in low SNRs.

The unweighted PVM, DF-III model was superior to the PVM, DF-II, even though the DF-III model used the video-only voicing classifier that had a near guessing level performance. This suggests that an alternative configuration for the PVM, DF model might be to have an audio-only voicing classifier, video-only place classifier, and use an audio-visual, fused estimate for the manner classification. However, given an adaptive fusion strategy there would be no need for such a decision to be made. Indeed, the noisy trained, neurally weighted PVM fusion results in an overall error reduction of 9.83% and an optimally competitive reduction of 20.36% for SNRs 10dB and less. Moreover, this model gives a 20dB gain at 1dB when compared to the performance of the audio-only classifier. That is, the performance of this PVM model at 1dB is equivalent to an audio-only performance at 21dB - a 20dB gain.

6. References

[1] T. Lewis and D. Powers, "Distinctive feature fusion for improved audio-visual phoneme recognition," in *The Eighth International Symposium on Signal Proocessing and Its Applications*, A. Bouzerdoum and A. Beghdadi, Eds. Sydney, Australia: IEEE, 2005.

[2] G. Potamianos, C. Neti, G. Gravier, A. Garg, and A. Senior, "Recent advances in the automatic recognition of audiovisual speech," *Proceedings of the IEEE*, vol. 91, no. 9, pp. 1306–1326, Sept 2003.

[3] B. Walden, R. Prosek, A. Montgomery, C. Scherr, and C. Jones, "Effect of training on the visual recognition of consonants," *Journal of Speech and Hearing Research*, vol. 20, pp. 130–145, 1977.

[4] R. Jakobson, C. Gunnar, M. Fant, and M. Halle, *Preliminaries to speech analysis; the distinctive features and their correlates*. Cambridge, Mass.: MIT Press, 1967.

[5] Q. Summerfield, "Some preliminaries to a comprehensive account of audio-visual speech perception," in *Hearing by Eye: The pyschology of lip-reading*, B. Dodd and R. Campbell, Eds. Hillsdale NJ: Lawrence Erlbaum Associates, 1987, pp. 3–52.

[6] P. Niyogi, E. Petejan, and J. Zhong, "Feature based representation for audio-visual speech recognition," in *Auditory-Visual Speech Processing (AVSP'99)*, 1999.

[7] F. Berthommier, "Audio-visual recognition of spectrally reduced speech," in *Proceedings of AVSP'01*, Aalborg, Denmark, September 2001.

[8] J. Benki, "Analysis of english nonsense syllable recognition in noise," *Phonetica*, vol. 60, pp. 129–157, 2003.

[9] T. W. Lewis, "Noise-robust audio-visual phoneme recognition," Ph.D. dissertation, Flinders Univeristy, 2005.

[10] R. Goecke and J. Millar, "The audio-video australian english speech data corpus AVOZES," in *Proceedings of the 8th International Conference on Spoken Language Processing ICSLP2004*, vol. III, Jeju, Korea, October 2004, pp. 2525–2528.

[11] D. Powers, "Recall and precision versus the bookmaker," in *International Conference on Cognitive Science*. University of New South Wales, Sydney, July 2003, pp. 529–534.