Appendix B: Effect of Target Size

In this Appendix, we examine the effect of target set size on the probability of miss, P_{miss}, and the probability of false alarm, P_{FA}. We will examine the effect for the speaker ID problem, but the same analysis will apply to other problems, such as keyword spotting and language ID. This effect has been well documented in the literature [Singer and Reynolds, 2004; Zigel and Wasserblat, 2006].

For this analysis, we shall assume that the target speaker set, S, is a random set of N target speakers and that each acceptance or rejection of a speaker is independent. A hit is defined as the correct identification of any of the speakers in the set S. So, a miss is the false rejection of a speaker who is in the set S as none of the speakers in that set. Similarly, a false alarm is defined as the false acceptance of a speaker not in the set S as one of the speakers in the set S. We will now derive the probability of a miss and of a false alarm for the set of N speakers in terms of the corresponding probabilities for a single target speaker. For simplicity, we will assume that each of the target speakers has the same miss and false alarm probability.

Let

P_{miss} = probability of a miss (false reject) for a single target speaker, and

P_{FA} = probability of a false alarm (false accept) for a single target speaker.

In other words, P_{miss} is the probability that a sample from the target speaker is rejected in favor of the set of all speakers outside of the target set, and P_{FA} is the probability that the system recognizes a sample from a speaker outside the target set as the target speaker. We then conclude that the probability of correctly recognizing a sample from one of the speakers outside the target set as ***not*** from the target speaker, is given by:

P(correct reject) = $1 - P_{FA}$ = probability of correctly recognizing a sample not from the target speaker as such.

For a miss in the set S to take place, the system must decide that a sample from our target speaker is neither from that speaker (false reject) nor is it from any of the other N-1 speakers in the set S (correct reject). Because of the independence assumption, the joint probability of the miss event is the product of the individual probabilities. The probability of falsely rejecting the target speaker is just P_{miss}. The probability of correctly rejecting each of the other N-1 speakers in the set S is $1 - P_{FA}$. Therefore, the probability of a miss in the set is just the product of the component probabilities:

$$P(\text{miss-set}) = P_{miss}\,(1 - P_{FA})^{N-1} \qquad (1)$$

Now, for a false alarm in the set to take place, a test sample from outside the target set must be falsely accepted as one of the target speakers. So, the P(false-alarm-set) is one minus the probability of being correctly rejected by all of the N target models. In other words,

$$P(\text{false-alarm-set}) = 1 - (1 - P_{FA})^{N} \qquad (2)$$

The table below shows an example with P_{miss} =5% and P_{FA} =1% for different values of set size N. As one would expect, P(miss-set) does decrease as N increases, but very slowly. However, we also see that P(false-alarm-set) increases almost linearly with N, because there are N ways to have a false alarm. That can be seen from equation (2), which for small P_{FA} and values of N less than $1/P_{FA}$, can be approximated by:

P(false-alarm-set) $\approx N * P_{FA}$, for $P_{FA} << 1$, and $N << 1/P_{FA}$. (3)

P(miss-set) could also be approximated by:

P(miss-set) $\approx P_{miss} [1 - (N-1) P_{FA}]$, for $P_{FA} << 1$, and $N << 1/P_{FA}$.

So, although P(miss-set) decreases as N increases, the effect of the decrease is not as dramatic as the relative increase in P(false-alarm-set). Our conclusion is that having a larger set of target speakers does make a test somewhat easier in terms of probability of miss, but it makes it much, much harder in terms of probability of false alarm.

N	P(miss)	P(FA)
1	5.0%	1.0%
5	4.8%	4.9%
10	4.6%	9.6%
25	3.9%	22.2%
100	1.8%	63.4%

Now, the above analysis made a number of simplifying assumptions, so the question is whether the above described effect occurs in practice or not. We first present an experimental result and then we give a different scenario of interest.

In a project on speaker ID sponsored by AFRL [Gish and Bellfield, 2004], we performed experiments with target sets of multiple speakers (N=15, 30, 45) using data from the TIMIT database, supplied to us by AFRL. The speaker ID experiments we performed showed that the experimental ROC curves for each of the three conditions very much followed the expected ROC curves as would be predicted from (1) and (2) above.

The independence assumptions made in the above analysis might not hold for certain applications. For example, there might be interest in a set of speakers who use the same communication channel, so you would expect those speakers to sound more similar to each other and, therefore, be more correlated. In keyword spotting, the set of keywords may be similar to each other, e.g., variations of pronunciations of the same word in different dialects. In such cases, one would not expect the false alarm rate to increase as rapidly with increased set size.

E. Singer and D. Reynolds, "Analysis of Multitarget Detection for Speaker and Language Recognition," *IEEE Odyssey: The Speaker and Language Recognition Workshop*, Toledo, OH, 2004.

Y. Zigel and M. Wasserblat, "How to Deal with Multiple Targets in Speaker Identification Systems," *IEEE Odyssy: The Speaker and Language Recognition Workshop*, pp. 1–7, 2006.

H. Gish and W. Bellfield, "Improved Methods for Open Set Speaker Identification," *Final Report*, AFRL Contract F30602-03-C-0096, 2004.