

Zooplots for Speaker Recognition with Tall and Fat Animals

Anil Alexander¹, Oscar Forth¹, John Nash², and Neil Yager³

¹Oxford Wave Research Ltd, ²University of York, ³AICBT Ltd, United Kingdom

{anil|oscar@oxfordwaveresearch.com, neil@aicbt.com}

Performance in speaker recognition is normally discussed using database-centric single figures of merit such as equal error rates. These metrics fail to capture the performances of individual speakers or speaker groups, which are very important in forensic speaker recognition. For instance, a recognition system that works well for male speakers may perform poorly for female speakers. Alternatively, a system may fail for speakers of a certain language or under a specific recording condition. The zoo-plot analysis, developed by Yager and Dunstone (2011), extends George Doddington's (1998) original classification of the biometric menagerie to categorise other difficult speakers. Under the original Doddington classification, *sheep*, who are 'normal' speakers and tend to match well against themselves and poorly against others, are the majority of the speakers within the database. *Goats* are speakers who are difficult to verify and tend to have low genuine match scores. *Lambs* generally match with high scores against other speakers and are thus easily impersonated, resulting in false accepts. *Wolves* easily impersonate other speakers, also resulting in false accepts. Yager and Dunstone extend this menagerie by taking both genuine and imposter performance into consideration, leading to four new 'animal' types: *chameleons*, *phantoms*, *doves* and *worms*. *Chameleons* always appear like others, receiving high scores for matches against themselves and others. *Phantoms* always receive low scores, so rarely match against themselves or others. *Doves* are the best possible users of a recognitions system, as they have high scores when matched against themselves and low scores when matched against others. *Worms* are the worst users of a biometric system, and are characterised by low genuine scores and high imposter scores.

Zooplot analysis is performed as follows: Select a group of speakers that represents a recording condition. From this set of speakers, select non-contemporaneous files for testing and training speakers. Ideally, there should be more than one file each for testing and training for the same speaker. For each speaker, match their training samples against all of their testing samples and compute their average genuine match score. Similarly, the mean of all the scores obtained by comparing his/her training samples with files from other speakers gives the average imposter score. In a two-dimensional quartile plot, as shown in Figure 2, the average genuine score is plotted against the average imposter score for all speakers. The users who fall within the four quartiles (top and bottom 25%) are assigned to the animal groups (*worms*, *chameleons*, *doves* and *phantoms*), with each set showing different characteristics.

In this work, we further extend the classification of these animals by characterising the speakers as 'tall/short' or 'fat/thin', depending on the variability of their genuine and imposter match scores (see Figure 3). For example, if a 'dove' speaker has low genuine variability and high imposter variability, then he or she is a 'tall thin dove'. Generally speaking, variability of match scores is symptomatic of an underlying problem, regardless of animal type. Therefore, the enhanced visualization adds a new dimension of independent and useful diagnostic information.

While single figures of merit like equal error rates provide information about performance of a system against a database as a whole, zooplot analysis can provide valuable insight into the properties of individual speakers and clusters of speakers in the database. It can help to identify potential algorithmic weaknesses of systems against certain classes of speakers, and can be used to adjust identification thresholds at an individual or group level. Preliminary research seems to suggest a link between certain aspects of voice quality and speaker categories in the zooplots. We recommend that zooplot analysis is done as speakers are added into a database, to help identify commonalities of speaker groups or algorithmic weaknesses of systems.

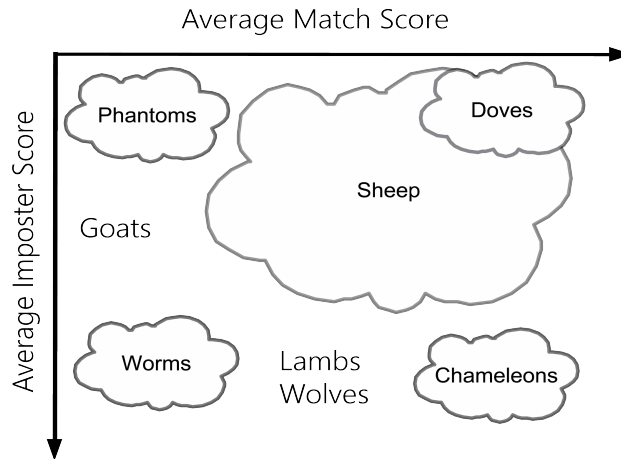


Figure 1: Illustration of a zooplots described in Yager and Dunstone 2011



Figure 2: Zooplot using speakers from the IPSCO3 database using the VOCALISE spectral comparison

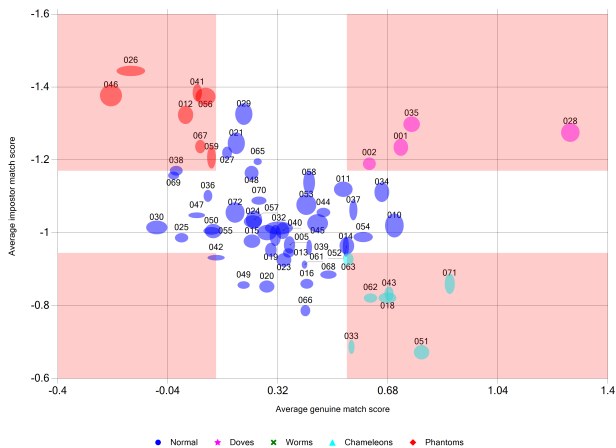


Figure 3: ‘Tall and Fat’ extension of Zooplot in Figure 2

References

- N. Yager, and T. Dunstone, Biometric Systems for Data Analysis: Design, Evaluation, and Data Mining, 2009 Springer Press, ISBN-13:978-0-387-77625-5
- G. Doddington, W. Liggett, A. Martin, M. Przybocki, D.Reynolds, Sheep, goats, lambs and wolves: a statistical analysis of speaker performance, Proceedings of IC-SLD'98, NIST 1998 Speaker Recognition Evaluation, Sydney, Australia, November 1998, pp. 1351–1354.