Conclusions on voice comparison evidence in Germany and a challenging case

20th IAFPA conference
Vienna, July 24-27, 2011

Michael Jessen
BKA, Department of Speaker Identification and Audio Analysis (KT54)
Conclusions on voice comparison evidence in Germany and a challenging case

Structure

Part I:
Conclusions on voice comparison evidence in Germany: the classical approach and ways to capture the Bayesian approach

Part II:
A problem with the classical approach: The brother challenge

Conclusion
Identity or non-identity
- cannot be given (non liquet)
- applies with predominant probability
- applies with high probability
- applies with very high probability
- applies with a probability close to certainty

In Bayesian terms, these are verbalized forms of posterior probabilities.
Ways to capture the Bayesian approach I

Köller et al. (2004) (handwriting experts)

Odds form of Bayes’ Theorem

\[
\frac{p(H_{so} | E)}{p(H_{do} | E)} = \frac{p(E | H_{so})}{p(E | H_{do})} \times \frac{p(H_{so})}{p(H_{do})}
\]

posterior odds  likelihood ratio  prior odds

SO=same origin; here same speaker
DO=different origin; here different speaker

Morrison (2010)

\[
\frac{p(H_{so} | E)}{p(H_{do} | E)} = \frac{p(E | H_{so})}{p(E | H_{do})} \times \frac{p(H_{so})}{p(H_{do})}
\]

Unfortunately, this method of using 1 as prior odds is also known as the Prosecutor’s Fallacy.

According to the “Indifference” principle (Köller et al. 2004) or the “Maximal Entropy” (maximal uncertainty) principle (Katterwe 2006)
The task: Comparing car paint left at the scene with car paint from a known truck (case at the BKA from 2003/04).

The analysis: The car paint composition was quite unique: multiple layers and contamination spots due to several over-paint actions.

The probability of obtaining this particular composition of car paint had it come from a different truck was estimated to be 1/100 million. It was assumed that the numerator of the LR was 1, so the LR was estimated to be 100 million.

The number of similar-size trucks registered in Germany (4.6 million) was taken for the prior odds, resulting in a value of 1/4.6 million.

The resulting posterior probability is about \( p = \frac{100}{100+4.6} \). (cf. Robertson & Vignaux 1995: 15 on how to turn from odds to probability)

Conclusion: With 95% probability, the car paint at the scene originates from the known truck, with 5% probability it has a different origin.
The case worked well (although 95% is still not spectacular).

The reason it worked was because the LR was extremely high and much higher than the number of existing entities for the different-origin hypothesis.

In speech analysis it is unrealistic to obtain LRs of that size. Speaker identification cannot learn from this case, because it can practically never proceed in an analogous fashion.

The problem with the case is that it strongly overestimates the number of entities (trucks) for the different-origin hypothesis. The number of entities *that could have realistically been involved based on other evidence and circumstances* is much smaller.

This overestimation is also known as the *Defence Attorney’s Fallacy*. 
Stage 1:
In a voice comparison case, the quality/quantity of the material and the constellations of similarity and typicality are such that the conclusion “with very high probability identical” is given.

Stage 2:
Later it turns out that the brother of the accused could have been the perpetrator. Therefore, a second voice comparison is ordered. The voice of brother B is similar to the voice of the brother A, yet not as close to the voice of the unknown as in the case of brother A. Without knowledge about brother A, the conclusion for brother B would be “with predominant probability identical”.

This is not possible logically, because the probabilities for both brothers would add up to a total probability larger than 1. This would be in violation of one of the Kolmogoroff axioms: p(E)=1, i.e. the probability of the event space is 1 (see Köller et al. 2004).
Possible numerical equivalents of verbal posterior probabilities

<table>
<thead>
<tr>
<th>Verbal probability</th>
<th>Numerical equivalent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-liquet</td>
<td>0.5</td>
</tr>
<tr>
<td>with predominant probability</td>
<td>0.75</td>
</tr>
<tr>
<td>with high probability</td>
<td>0.90</td>
</tr>
<tr>
<td>with very high probability</td>
<td>0.95</td>
</tr>
<tr>
<td>with a probability close to certainty</td>
<td>0.99</td>
</tr>
</tbody>
</table>
Illustration of the Kolmogoroff violation and the necessity for readjustment

The problem: there is no way to fit in another 0.75 ("predominant probability") for Brother B into the 0.05 slot. So all the probabilities have to be readjusted, e.g.:

Stage 1

- \( p(H_{\text{Brother A}|E}) = 0.95 \)
- \( p(H_{\text{someone else}|E}) = 0.05 \)

Stage 2

- \( p(H_{\text{Brother B}|E}) = 0.125 \)
- \( p(H_{\text{someone else}|E}) = 0.125 \)
- \( p(H_{\text{Brother A}|E}) = 0.75 \)
Likelihood Ratio alternative

No such problems and no need for readjustment occur with the Likelihood Ratio (LR) or its verbal equivalents, e.g.

Brother A: LR = 120  Brother B: LR = 20

The fact that both LRs would support identity is not a logical problem (but still a practical one). Prior odds and further evidence might show that the posterior odds are much higher for one brother than the other.
Calculations in the Bayesian framework are **binary**, with exclusive and (usually) exhaustive hypotheses (Robertson & Vignaux 1995).

The ternary situation (Brother A, Brother B, other) is outside of this framework. Therefore, Kolmogoroff’sian space only needs to accommodate the probabilities of two hypotheses.

This would mean that it is logically possible to have $p=0.95$ for Brother A and $p=0.75$ for Brother B.

This is still odd though, and difficult to justify in court. It suggests that these posterior probabilities were too high in the first place.* But making them much lower would render them useless.

Reporting Likelihood Ratios instead would not create these types of unconvincing effects.

*In order to get these posterior probabilities with the LRs shown above, the prior odds have to be around 6, which is pretty low.
The BKA (or the local LKÄ or private experts in Germany) report classical probabilities. Within the Bayesian approach this means to report **posterior odds** (or posterior probabilities, which is equivalent).

Posterior odds require knowledge of the Likelihood Ratio plus the prior odds, but the **prior odds are not known** to the speech expert.

Attempts have been made in Germany to solve this problem, but they have turned out to be versions of the Prosecutor’s Fallacy or the Defence Attorney’s Fallacy.

The **brother challenge** further highlights problems with the classical approach. But unlike previously thought (and found in the abstract booklet), it might not create a logical problem.

The currently most convincing position remains that the **speech expert should not report posterior odds, but only Likelihood Ratios or verbalizations of the Likelihood Ratio concept** (Rose 2002, Rose & Morrison 2009, French et al. 2010, Morrison 2010).


