VOICE BIOMETRICS WITHIN THE FAMILY: TRUST, PRIVACY AND PERSONALISATION

Delphine Charlet
France Telecom
2 avenue Pierre Marzin, 22307 Lannion Cedex, France

Victor Peral Lecha
France Telecom
Building 10, Chiswick Business Park, 566 Chiswick High Road, London W4 5XS, United-Kingdom

Keywords: Voice biometrics, speaker recognition, privacy, trust, personalisation, family voices evaluation.

Abstract: Driven by an increasing need for personalising and protecting access to voice services, the France Telecom R&D speaker recognition system has been used as a framework for experimenting with voices in a family context. With the aim of evaluating this task, 33 families were recruited. Particular attention was given to 2 main scenarios: a name-based and a common sentence-based scenario. In each case, members of the family pronounce their complete name or a common sentence respectively. Moreover, this paper presents a database collection and first experiments for family voices.

The results of this preliminary study show the particular difficulties of speaker recognition within the family, depending on the scenario, the genre and age of the speaker, and the physiological nature of the impersonation.

1 INTRODUCTION

Privacy within a family is a sensible subject: each member of a family may require some privacy, but achieving privacy protection through traditional means (such as pin code) may be resented by the other members of the family. Indeed, traditional privacy protection is based on “what you know” information, thus, to be effective, it needs to be hidden from the others. So the use of traditional authentication password or PIN is inadequate with the global feeling of mutual trust that often exists within the family. Hence, speaker recognition appears to be an interesting way for privacy protection since it is based on “what-you-are” and not on “what-you-hide”. On the other hand, the genetic links that exist between parents and children as well as the same cultural and geographical contexts they share may hamper speaker recognition, as it is pointed by (van Leeuwen, 2003). Moreover, children constitute a difficult target for speech recognition in general (Wilpon and Jacobsen, 1996).

As far as we know, no published study has been done on this subject. Thus, for the purpose of this preliminary study we collected a database of family voices in order to perform technological evaluation of the performances of speaker recognition in the context of family voices.

2 APPLICATION DESIGN

We have defined a standard application to evaluate the feasability to personalise home telecoms services for each member of a family, based on voice recognition. It has been decided that personalisation should be done through an explicit step of identification/authentication. Thus, if this step is explicit, it has to be as short as possible. These ergonomic considerations lead us to two technological choices:

- Identification and authentication performed on the same acoustical utterance.
- Text-dependent speaker recognition: as it achieves acceptable performances for short utterances (less than 2 seconds).

Two scenarios were defined, depending on the content of the utterance pronounced by the speaker.

2.1 Name-based scenario

In this scenario, the user pronounces his/her first and last name to be identified and authenticated. The main characteristics of such an application are:

- It is easy to recall.
- It is usually rather short.
- It enables deliberate impostor attempts: e.g. a mother can claim to be her daughter.

Charlet D. and Peral Lecha V. (2005).
VOICE BIOMETRICS WITHIN THE FAMILY: TRUST, PRIVACY AND PERSONALISATION.
In *Proceedings of the Second International Conference on e-Business and Telecommunication Networks*, pages 330-334
DOI: 10.5220/0001407003300334

As for each member of the family, the phonetic content of the voiceprint is different, the identification process is performed on both the voice and the phonetic context, although there is no explicit name recognition process. Thus it can be expected that the identification error rate will be negligeable. On the other hand, as the authentication is performed on a short utterance (e.g. 4 syllables), the authentication performances are expected to be rather poor.

2.2 Common sentence-based scenario

In this scenario, the user pronounces a sentence, which is common to all the members of the family. The main characteristics of such an application, compared to the name-based scenario are:

- It is less easy to recall: the sentence should be prompted to the user, if he does not remember.
- The length of the sentence may be longer than a simple name.
- It prevents deliberate impostor attempts: e.g. a mother can not claim to be her daughter.

As for each member of the family the phonetic content of the voiceprint is the same, the identification process is performed only on the voice. Thus it can be expected that the identification error rate will be higher than in the context of name-based recognition. On the other hand, as the authentication is performed on a longer sentence (10 syllables), the authentication performances are expected to be better than those of the name-based scenario.

3 DATABASE DESIGN

3.1 Family profile

The families were required to be composed of 2 parents and 2 children. The children are older than 10. They all live in the area of Lannion, where this work was conducted. 33 families were recruited: 19 with one son and one daughter, 10 with 2 sons and 4 with 2 daughters. They were asked to perform 10 calls from home, with their landline phone, during a period of one month. Hence, factors such as voice evolution over time or sensitivity to call conditions are not studied in this work.

3.2 Name-based scenario

In this scenario, the key point is the fact that there might be deliberate impostor attempts on a target speaker, as the user claims an identity to be recognised.

3.2.1 Training

For each member of a family, training is performed with 3 repetitions of his/her complete name (first + last name). This number of repetitions represents a good trade-off between performances and tediousness of the task.

3.2.2 Within family attempts

Each member of a family is asked to perform attempts on his own name, and also attempts on the name of each of the other members of his family.

3.2.3 External impostor attempts

A set of impostor attempts from people who do not belong to the family is collected. The external impostors are composed of members of other families who pronounced the name of the member of the target family.

3.2.4 Collected Database

As we focus on speaker recognition, we only retained the utterances where the complete name was correctly pronounced.

- 16 families completed their training phase.
- 13 families completed the testing phase also.
- 672 true speaker attempts collected for the 13 families, that makes an average of 52 true speaker attempts per family, thus an average of 13 true speaker attempts per user.
- 582 within family impostor attempts collected for the 13 families, that makes an average of 45 within family impostor attempts per family, thus an average of 11 within family impostor attempts per user.
- 2173 external impostor attempts (impostor attempts are performed on all the families who have completed the training phase)

3.3 Common sentence-based scenario

In this scenario, the key point is the fact that there can not be deliberate impostor attempts on a target speaker, as the user does not claim identity to be recognised.

3.3.1 Training

For each member of a family, training is performed with 3 repetitions of the common sentence.

3.3.2 Within family attempts

Each member performs attempts by pronouncing the common sentence.

3.3.3 External impostor attempts

The attempts of other families are used to perform external impostor attempts.

3.3.4 Collected Database

As we focus on speaker recognition, we only retain the utterances where the sentence was correctly pronounced.

- 25 families completed their training phase.
- 17 families completed the testing phase also.
- 614 within family attempts collected for the 17 families, that makes an average of 36 attempts per family, thus an average of 9 attempts per user.
- 13549 external impostor attempts: all within family attempts are used to perform impostor attempts on the other families (impostor attempts are performed on all the families who have completed the training phase).

4 EXPERIMENTS

4.1 Speaker recognition system

For these experiments, the text-dependent speaker recognition system developed in France Telecom R&D is used. It basically relies on HMM modelling on cepstral features, with a special care on variances, with a speaker-independant contextual phone loop as reject model (Charlet et al., 1997). The task is open-set speaker identification. There was no particular tuning of the system to this new task. The rejection thresholds are set a posteriori, common to all families.

4.2 Typology of errors

4.2.1 Name-based scenario

In this scenario, the identity claim is taken into account in the typology of errors. Within the family, when the speaker A claims to be speaker A (pronouncing the name of speaker A), the system can:

- accept the speaker as being speaker A: correct acceptance (CA)
- accept the speaker as being speaker B: false identification error (FI)
- reject the speaker: false rejection error (FR)

Moreover, in this scenario, within the family, a speaker can make deliberate impostor attempts on a target speaker. When the speaker A claims to be speaker B, the system can:

- reject the speaker: correct rejection on internal impersonation (CRII)
- accept the speaker as being speaker B: wanted false acceptance on claimed speaker (WFA)
- accept the speaker as being speaker A: unwanted correct acceptance (UCA)
- accept the speaker as being speaker C: unwanted false acceptance (UFA)
- Internal False Acceptance is defined as: IFA = WFA + UFA

Outside the family, when an impostor claims to be speaker A, the system can:

- reject the speaker: correct external rejection (CER)
- accept the speaker as being speaker A: external false acceptance on claimed speaker (EWFA)
- accept the speaker as being speaker B: external unwanted false acceptance (EUFA)
- External False Acceptance is defined as: EFA = EWFA + UFA

4.2.2 Common sentence-based scenario

This is the classical typology of open-set speaker identification. Within the family, when the speaker A pronounces the common sentence, the system can:

- accept the speaker as being speaker A: correct acceptance (CA)
- accept the speaker as being speaker B: false identification error (FI)
- reject the speaker : false rejection error (FR)

Outside the family, when an impostor pronounces the common sentence, the system can:

- accept the speaker as being speaker A: external false acceptance error (EFA)
- reject the speaker: correct external rejection (CER)

5 RESULTS

5.1 name-based scenario

Figure 1 plots EFA, IFA and FI as a function of FR for the name-based system.

For the particular operating point where { FR=7.9%; FI=0.4%; IFA=21.1%; EFA=8.1%}, let us see the details of errors:

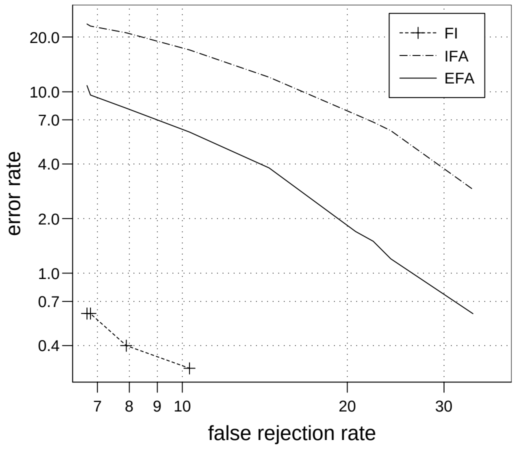

Figure 1: Name-based speaker identification within the family

- FI=0.4% is low and due to 2 types of surprising errors: a man identified as his wife and a 10-year-old girl identified as his father.
- UCA=20.6% : when a member within the family claims to be another member, he is "unmasked" (i.e. truly identified) in 20% of the cases. This is mainly due to the fact that, as the voiceprint of the members of the family shares important part of the phonetic content (same last name), an attempt with another first name may match reasonably well the voiceprint of the speaker if the first names are not too different.
- IFA=21.1% - EFA=8.1% : the rate of IFA (false acceptance from deliberate impostor attempts within the family) is more than 2.5 times as much as EFA (from deliberate impostor outside the family, with the same age/gender distribution as the impostors within the family). Although we cannot quantify what is due to line effects from what is due to physiological resemblance, we can note that the system is much more fragile to impostor within the same family than from outside the family.

IFA is decomposed into WFA=18.0% and UFA=3.1%. Table 1 presents the false acceptance rate for deliberate impersonation (WFA : when speaker A claims the name of speaker B and is accepted as being speaker B) within the family, for the different types of speaker. "Son" can do impostor attempts on "son" when they are brothers. In the same way, "daughter" can do impostor attempts on "daughter" when they are sisters.

We analyze each type of speaker (father, mother, daughter, son) with respect to their ability to defeat the system and to their "fragility" to impostor attempts (Doddington et al., 1998). In the table, between brackets are given the numbers of tests of each type of imposture, in order to give an idea of the reliability of the results.

Table 1: false alarm rate according to the type of impostor and target speaker

Impostor	target speaker			
speaker	father	mother	son	daughter
father	–	11.8 [51]	12.3 [57]	2.2 [45]
mother	0 [52]	–	14.8 [54]	18.6 [43]
son	7.4 [54]	29.8 [47]	29.4 [17]	53.3 [30]
daughter	0 [45]	32.6 [43]	50 [36]	50 [8]

From the table, the following remarks can be done:

- the father, as an impostor, gets a moderate and equal success rate on his wife and his son, and a low success on his daughter. As a target, he is "fragile" against his son at a moderate level.
- the mother, as an impostor, gets a better success rate than her husband on her children, and a complete failure on her husband. The difference of success rate between the son target and the daughter target is not high. As a target, she is fragile against her husband at a moderate level, and she is equally fragile at a high level against her children.
- the son, as an impostor, has a low success rate on his father, a high level on his mother and brother and a very high level on his sister (the difference between brother and sister success rate might not be significant because of the small number of attempts in the case of 2 brothers attempts). As a target, he is moderately fragile against his parents and highly against his sister.
- the daughter, as an impostor, has a high success rate on her mother and a very high success on her sister or brother, and a complete failure on her father. As a target, she is not fragile against her father, moderately against her mother and very highly against her sister or brother.

5.2 common sentence-based scenario

Figure 2 shows the global performances of the common sentence-based system, for different operating points.

For a particular operating point where globally {FR=6.1%; FI=1.2%; EFA=5.8%;}, analyzing the re-

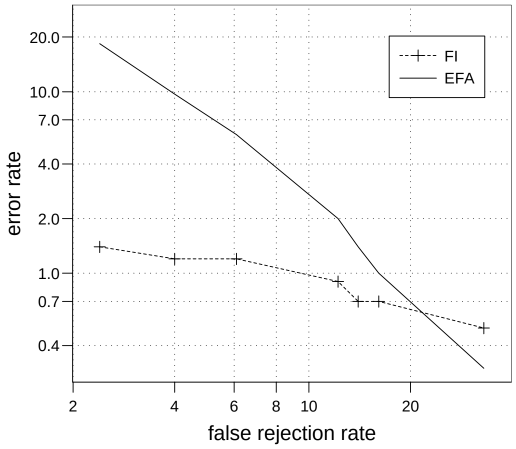

Figure 2: Common sentence-based speaker identification within the family

sults family per family, we observe that over the 17 families who made identification attempts:

- 13 families have no false identification : FI=0%
- 3 families get a false identification rate : FI=2-2.5% (actually 1 observed error)
  - family #13: father identified as his 17-year-old son
  - family #16: 14-year-old girl identified as her mother
  - family #20: mother identified as her 20-year-old girl
- 1 family get a false identification rate of FI=7% due to one unique cause of error: a 12-year-old girl was identified as her mother

Speaker identification within a family appears to be easily achievable for a majority of the families tested (13/17). For the remaining families, where errors were observed, the errors are not frequent and always concern a specific pair parent/same-sex teenager.

6 DISCUSSION

In the name-based scenario, there is deliberate impersonation within the family. As the speech utterance to perform speaker authentication on is very short (on average 4 syllables) performances are very poor. However, the analysis of the differences of performances according to the type of impostor and target speaker is interesting. It shows that the most “fragile” to impersonation are the children, between them, and that they are also the most effective impostors.

In the common sentence-based scenario, there cannot be deliberate impersonation, as there is no identiy claim. Except for some rare cases where there is an identification error, it seems to be the most effective and ergonomic way to perform speaker recognition within the family. Considering the case of identification errors, as they always occured on the same pair, it should be possible to develop special training procedure to prevent them.

7 CONCLUSION

In this paper, we have presented a database of family voices collected for voice recognition within the family. Two scenarios are studied. The number of recorded attempts is enough to perform experiments and draw first conclusions about the particular difficulties of speaker recognition within the family.

Despite the novel nature of the application, voice biometrics seems the most natural way to manage the privacy and trust issues when accessing to voice services in a family context. Further research will attempt to improve the reliability of the system when dealing with the voices of children.

REFERENCES

Charlet, D., Jouvet, D., and Collin, O. (1997). Speaker verification with user-selectable password. In *Proc. of the COST 250 Workshop*, Rhodos, Greece.

Doddington, G., Liggett, W., Martin, A., Przybocki, M., and Reynolds, D. (1998). Sheeps, goats, lambs and wolves, a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *Proc. of the ICSLP 1998*, Sydney, Australia.

van Leeuwen, D. (2003). Speaker verification systems and security considerations. In *Proc. of Eurospeech 2003*, pages 1661–1664, Geneva, Switzerland.

Wilpon, J. and Jacobsen, C. (1996). A study of speech recognition for children and the elderly. In *Proc. of the ICASSP 1996*, pages 349–352, Atlanta, USA.