

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月28日現在

機関番号：82505

研究種目：基盤研究（C）

研究期間：2009～2012

課題番号：21510185

研究課題名（和文） ベイズ統計に基づく話者の異同識別鑑定における尤度比尺度の改良

研究課題名（英文） Improvement of likelihood ratio measurement in a forensic speaker identification based on Bayesian statistics

研究代表者

長内 隆（OSANAI TAKASHI）

科学警察研究所・法科学第四部・室長

研究者番号：70392264

研究成果の概要（和文）：

法科学分野においては、裁判官らによる適切な判断を助けるために、被疑者が犯人である可能性の程度を示すことが重要である。この可能性を示すために、ベイズ統計に基づく尤度比が広く利用されている。近年、この尤度比を話者の異同識別鑑定に利用する研究が行われている。しかし、従来法では、与えられた音声資料の一部しか利用できない。本研究では、与えられた音声資料を有効に利用できる尤度比尺度を提案し、その有効性を確認した。

研究成果の概要（英文）：

In the forensic science field, in order to help the suitable judgment by judges, it is important to show the degree of a possibility that a suspected person is a criminal. In order to show this possibility, the likelihood ratio based on Bayesian statistics is used widely. In recent years, research which uses this likelihood ratio for forensic speaker recognition is carried out. However, by the conventional method, only a part of the given speech data is used. In this study, I proposed the likelihood ratio measurement which can be used without making useless the given speech data, and confirmed the effectiveness of using it.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	900,000	0	900,000
2010年度	1,000,000	0	1,000,000
2011年度	800,000	0	800,000
2012年度	600,000	0	600,000
年度			
総計	3,300,000	0	3,300,000

研究分野：複合新領域

科研費の分科・細目：社会・安全システム科学、社会システム工学・安全システム

キーワード：統計数学、音声学、ベイズ統計、話者認識、犯罪捜査支援

1. 研究開始当初の背景

一般の話者認識では、未知話者の音声は既知話者の音声であるか否かを判断し、その結果は、正解、不正解のいずれかで表される。一方、法廷での利用を前提とした法科学分野

における話者の異同識別鑑定では、犯人の音声と被疑者の音声は同一人の音声か否かを判断する際、例えば、「同一人の音声である可能性は90%である」といったように、被疑者が犯人である可能性の程度を示すことが

望まれる。法廷における最終判断は、さまざまな証拠に基づいて裁判官や国民から選ばれる裁判員が下すものであり、このような可能性を示すことは、裁判官、裁判員が適切な判断を行うのを助ける効果がある。

犯人と被疑者の音声から、それらの発声者が同一話者であるかどうかの可能性を示すということは、音声資料が与えられたことによって、被疑者が犯人である可能性がどのように変化するかを示すことと等価である。このような推論には、ベイズ統計学の利用が適している。ベイズ統計学とは、条件付き確率に基づいた統計学であり、事後確率は、事前確率と観察によって得られる尤度比の積で表されるという考えである。この事後確率を得るためには、事前確率が必要となるが、実際の事件においてはそれを知ることはできないことが少なくない。そこで法科学分野における話者の異同識別鑑定では、事後確率を示す代わりに、尤度比を提示することにとどめることが多い。この尤度比を利用した判断は、親子鑑定やDNA型鑑定などの法医学の分野では既に用いられている。

近年、この尤度比を話者の異同識別鑑定に利用する研究が行われている。ここで扱う尤度比は、比較する音声資料が同一人の発声によるものであると仮定したときに得られる尤度を、別人であると仮定したときに得られる尤度で割った尤度比であり、この尤度比が1より大きければ同一人、小さければ別人と判断するものである。現在提唱されている尤度比尺度では、比較対象とする音素（主として定常母音）区間を比較する音声資料から複数個所見つけ出し、その音素区間の中央付近の1フレームの音響特徴量（フォルマント周波数など）を用いて尤度比を求めている。法科学分野で扱う音声資料は、品質が悪く、十分な発話長の音声を得られるとは限らない。また発話内容を指定することもできない。したがって、与えられた音声資料をできるだけ有効に利用できることが望ましい。しかし、既存の尤度比尺度では音声資料から当該音素区間の1フレームだけを対象とするものであり、必ずしも有効利用が図られていない。

2. 研究の目的

法科学分野における話者の異同識別を行う上で、必ずしも十分な発話量や音質の音声資料が得られるわけではない。また既存の尤度比尺度では、当該音素を複数個所抽出するものであり、当該音素が1か所しかない場合は、比較することすらできない。したがって、与えられた音声資料をできるだけ有効に利用できる比較方法を見出す必要がある。そこで本研究では、当該音素区間全体が利用できる尤度比尺度を提案し、その有効性を様々な観点から検証することを目的とする。音声資料を効率的に利用できる比較手法が開発で

きれば、認証性能の改善が期待できると考える。

3. 研究の方法

(1) 尤度比尺度の改良

当該音素区間に対応できる、つまり連続音声に対応できる尤度比尺度の定式化を行う。

既存の尤度比尺度を(1)式に示す。(D. V. Lindley, "A problem in forensic science", *Biometrika*, vol. 64, no. 2, pp. 207-213, 1997.) この式は、比較する音声資料が同一人の発声によるものであると仮定したときに得られる尤度を、別人であると仮定したときに得られる尤度で割った尤度比であり、この値が1より大きければ同一人、小さければ別人と判断する。

$$LR \approx \frac{\tau}{a\sigma} \times \exp\left[-\frac{(\bar{x} - \bar{y})^2}{2a^2\sigma^2}\right] \times \exp\left[-\frac{(w - \mu)^2}{2\tau^2} + \frac{(z - \mu)^2}{\tau^2}\right] \quad (1)$$

ここで、 m, n, \bar{x}, \bar{y} および σ_x^2, σ_y^2 は、比較する資料から抽出した当該音素の数、音素区間内の1フレームから求めた平均値および標準偏差であり、 μ および τ^2 は、母集団の平均値および分散である。また、 z, w, a は、 $z = \frac{\bar{x} + \bar{y}}{2}$, $w = \frac{m\bar{x} + n\bar{y}}{m+n}$, $a = \sqrt{\frac{1}{m} + \frac{1}{n}}$ で表される。また、 σ^2 は、

$$\sigma^2 = \frac{(m-1)\sigma_x^2 + (n-1)\sigma_y^2}{m+n-2}$$

とする。

この式では、比較する2つの資料の平均値 \bar{x}, \bar{y} は、それぞれ以下の正規分布に従うものと仮定している。

$$\bar{x} \sim N(\theta_x, \sigma^2/m),$$

$$\bar{y} \sim N(\theta_y, \sigma^2/n).$$

この \bar{x}, \bar{y} は、 m または n 個の区間内の1フレームの音響特徴量から算出する。したがって、1区間を対象とした場合は1フレームしか音響特徴量が得られないため、 σ^2 を求めることができない。

そこで、比較する2つの資料の平均値 \bar{x}, \bar{y} は各々の資料の標本分散に従って生起するものと仮定する。つまり、

$$\bar{x} \sim N(\theta_x, \sigma_x^2),$$

$$\bar{y} \sim N(\theta_y, \sigma_y^2).$$

とみなす。この仮定に基づいて尤度比を算出すると、(2)式ようになる。 σ_x^2, σ_y^2 は、 m

または n 個の区間内の連続する数フレームから算出するので、1 区間を対象とした場合であっても求めることができる。以下で、この提案式を用いたときの有効性を検討した。

$$LR \approx \frac{\tau}{\sqrt{\sigma_x^2 + \sigma_y^2}} \times \exp\left[-\frac{(\bar{x} - \bar{y})^2}{2(\sigma_x^2 + \sigma_y^2)}\right] \quad (2)$$

$$\times \exp\left[-\frac{(w - \mu)^2}{2\tau^2} + \frac{(z - \mu)^2}{\tau^2}\right]$$

(2) 評価用テストセットの構築

法科学分野では、さまざまな環境下で収録された音声を対象とする。したがって、法科学分野での実用化を考えた場合、各々の環境下で性能を評価することが望ましい。そこで、法科学用の話者認識システムの性能評価を行うためのテストセットを構築について検討する。

(3) 音声データベース

科学警察研究所では、これまでに数種類の伝送系を介して収録した多数話者音声データベースを構築している。ここではそれらのうち、以下の3つの音声データベースを用いて検証を行った。

- ・第1世代の音声データベース (G1) : ダイアル式の600型黒電話を介して時期差約3~4か月で2時期に亘って収録した音声データベース。ここでは男性360名の母音発話を用いた。サンプリング周波数10kHz、量子化精度12bitの音声データをフレーム長25.6msec、フレーム周期12.8msec、ハミング窓をかけて、12次のLPCケプストラム係数を求めた。
- ・第2世代の音声データベース (G2) : ボタン式の多機能電話を介して時期差約2~3か月で3時期に亘って収録した音声データベース。ここでは男性300名の母音発話を用いた。サンプリング周波数11.025kHz、量子化精度16bitの音声データをフレーム長25.6msec、フレーム周期12.8msec、ハミング窓をかけて、12次のLPCケプストラム係数を求めた。
- ・第3世代の音声データベース (G3) : コンデンサマイクロホンを通じて時期差約2~3か月で2時期に亘って収録した音声データベース。ここでは男性50名の日本語音韻発話を用いた。サンプリング周波数44.1kHz、量子化精度16bitの音声データを8kHzにダウンサンプリングし、フレーム長6.4msec、フレーム周期1.6msec、ハミング窓をかけて、12次のLPCケプストラム

ム係数を求めた。

4. 研究成果

(1) 提案した尤度比尺度の性能評価

① 1発話を対象とした話者照合

提案した尤度比尺度の有効性を評価するために、音声データベースG1中の単独発声5母音を用いて話者照合実験を行った。なお、照合および母集団の統計量の算出に360名の話者を用いた。

既存の尤度比尺度を利用するにあたっては、連続するフレームを抽出した複数個のフレームとみなして尤度比を算出した。また、比較のため、当該母音区間の音響特徴量の平均値間のユークリッド距離を用いた照合実験も行った。その結果を図1に示す。ここでは、5母音の平均話者照合率を示す。

既存の尤度比尺度を用いた場合、連続するフレームから抽出したとみなしていることが分散の過小評価につながり、そのため、ユークリッド距離を用いた照合率よりも低くなるといった傾向が見られる。一方、提案した尤度比尺度の利用によって、いずれの母音においても照合率の向上が認められた。

② 複数発話を対象とした話者照合

既存の尤度比尺度は、元来、複数発話から抽出した資料の比較に用いるものである。そこで、複数発話を対象とした話者照合実験を行った。ここでは、1時期に3発話発声している音声データベースG2中の単独発声5母音を用いた。なお、母集団の統計量を算出は、比較する音声資料の話者以外の話者を用いて求めた。照合実験は300名で行った。比較に用いる発話数を変化させたときの話者照合率を図2に示す。ここでは、5母音の平均話者照合率を示す。

いずれの尺度においても、発話数の増加に伴って照合率が向上することがわかる。発話数が2または3の場合、提案法の照合率が高いことがわかる。ただしその差は発話数の増加とともに小さくなる傾向が見られる。本実験では3発話までしか比較できなかったが、

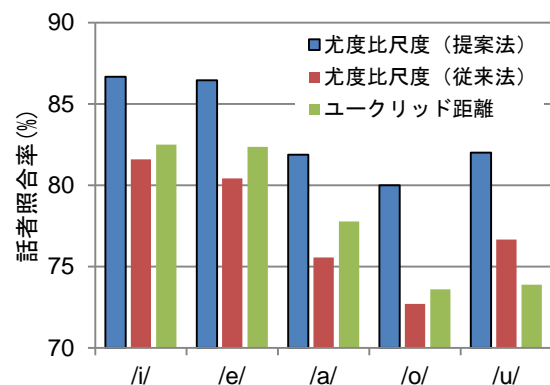


図1 尤度比尺度の違いによる話者照合率

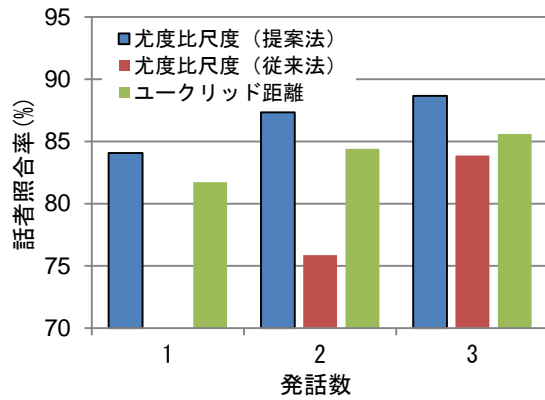


図2 照合に用いる発話数と照合率の関係

発話数をさらに増やした場合、提案法と従来法の照合性能差は縮まる可能性がある。なお、従来法では、発話数が1つの場合に尤度比を算出することはできない。この点は提案法の最大の利点である。

③ 複数母音の組み合わせ

これまでは1母音を対象とした話者照合実験を行ったが、複数の母音を組み合わせることで照合を行う効果について検討する。母音毎に提案法に基づく尤度比（対数尤度比）を求め、それらの和をとることで複数母音の尤度比を求め、照合を行った。使用した音声データベースはG2である。照合に使用する母音数を1～5に変化させたときの全組み合わせの話者照合率の平均を図3に示す。

照合に用いる母音数を増やすことで照合率は向上する。異なる3つの母音から1発話ずつ利用したときの照合率(91.8%)は、1母音で3発話用いたときの照合率(88.3%)よりも高くなることが分かった。これは話者照合を行う上で、音韻のバリエーションが多いほうが望ましいことを示唆するものである。

(2) 日本語音素を用いた話者照合

これまでは母音を用いてきたが、母音以外にも話者性情報を持つと言われている。そこで、日本語音韻を子音部と母音部に区分し、

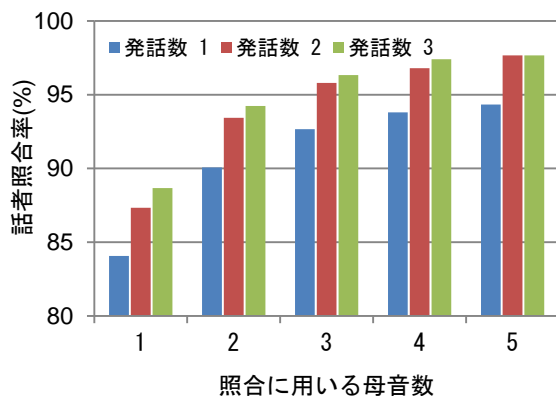


図3 照合に用いる母音数と照合率の関係

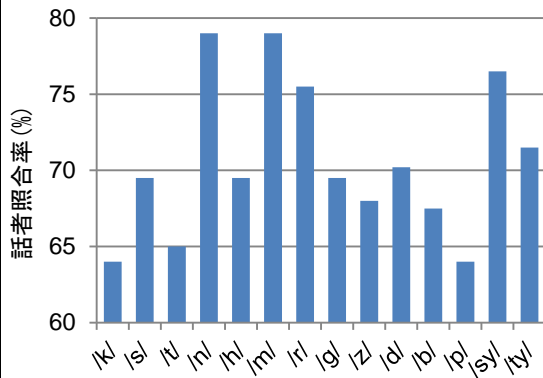


図4 日本語音韻の子音の話者照合率

各々の子音を用いて提案法に基づく尤度比（対数尤度比）による話者照合実験を行った。ここでは、音声データベース G3 中の単独発声した日本語音韻を用いた。なお、母集団の統計量を算出は、比較する音声資料の話者以外の話者を用いて求め、照合実験は50名で行った。照合に用いる子音の数は、後続母音が5つある場合は5、3つある場合は3とした。日本語音韻の各子音を用いたときの話者照合率を図4に示す。

子音/n/, /m/における照合率が比較的高い。これらの音は一般に、鼻音と呼ばれるもので、鼻腔の影響を受けることから話者性情報を多く含むと言われている。この照合実験結果からも、その点を裏付ける結果が得られた。鼻音と並んで高い照合率が得られた音素は/sh/である。これは摩擦音であり、先行研究でも高い照合性能を得ていることを確認しており、それを裏付ける結果となった。

(3) まとめ

尤度比を算出するため尺度を提案し、その有効性を確認した。特に、1発話しか得られないときは従来法では比較不可能であったが、提案法を用いることで比較可能となった。これは、比較に用いるのに十分な音声資料が得られにくい法科学分野における利用にあっては有利な点となる。なお、1発話だけでは高い識別性能を得ることは難しい。しかし、識別実験によって信頼性を示すことができるので、法廷における利用価値はあると考える。

(4) その他の識別実験

尤度比尺度の新たな利用として、母語識別、性別識別に適用できるか否かについて予備的検討を行った。母語識別については、日本語、中国語、韓国語、タイ語を母語に持つ話者による日本語音声を用いて、数字発話のイントネーションの違いや調音速度を特徴量とした日本語母語者と非母語者との識別実験を行い、90%以上の正答率を得た。性別識別については、基本周波数の利用が最も高い

識別率が得られること、フォルマント周波数の利用も効果的であること、調音速度などの時間領域の特徴量を併用することで性能の改善が図れることなどを確認した。いずれの識別実験も、オーソドックスな比較方法について検討し、その効果を確認した。今後はこれらの結果をベースラインとして、尤度比尺度を利用した識別結果との比較を行う予定である。

(5) 評価用テストセットの構築

科学警察研究所では、収録環境に対応できるようにするために、これまでダイアル式黒電話、固定電話、携帯電話を介した音声データベースを構築してきた。これらの音声データベースを用いることで、判断基準やその信頼性を示すことが可能となる。最近は、スマートフォン、IP電話などの新たな通信機器の普及や通信環境の整備が図られ、さらにLINEなどの通話アプリを利用した次世代の通信環境が浸透している。しかし、様々な環境下の音声資料をその都度用意して検証を行うことは困難であるとともに非効率的である。この点に適応するために、必要とする環境を構築し、疑似的な音声資料に変換する方法が考えられる。本研究では、このような疑似的な変換について試みたものの、検証実験を行うまでには至らなかった。今後は、この点についても取り組む必要があると考える。

(6) 今後の展望

本研究では、音声資料をできるだけ有効に利用できる尤度比尺度の提案を行い、その有効性を検証した。特に本提案法は、当該音素が1か所しかない場合でも比較可能な点に特長がある。ただし、個々の発話から尤度比を求めて後に結合するほうが、複数発話から尤度比を求める場合よりも照合率は低下することから、同じ音韻については、比較可能な複数発話を用いて尤度比を求めるほうが良い。また、比較できる音韻の数も多いほど認識性能は向上する。なお本検証では、単独に発声した日本語音韻中の母音や子音を用いたが、今後は、単語や文から比較対象となる音韻を抽出し、同様の検証実験を行った上で実用化に向けた検討を行うことが望まれる。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計3件)

- ① 長内隆、石原俊一、法科学分野における話者認識の動向、日本音響学会誌、査読無、69(7)、2013、ページ未定
- ② Kanae Amino, Takashi Osanai, Speaker Identification using Japanese

Monosyllables and Contributions of Nasal Consonants and Vowels to Identification Accuracy、Japanese Journal of Forensic Science and Technology、査読有、18、2013、13-21

- ③ Kanae Amino, Takashi Osanai, Speaker characteristics that appear in vowel nasalisation and their change over time、Acoustical Science and Technology、査読有、33、2012、96-105

[学会発表] (計19件)

- ① 網野加苗、蒔苗久則、鎌田敏明、長内隆、母音の無声化頻度と話者の出身地に関する考察、日本音響学会 2013 年春季研究発表会、2013 年 3 月 15 日、八王子市
- ② 網野加苗、長内隆、調音速度を用いた母語識別における発話内容の影響、日本音響学会、2013 年春季研究発表会、2013 年 3 月 13 日、東京都八王子市
- ③ 網野加苗、石原俊一、小川哲司、長内隆、黒岩眞吾、越仲孝文、篠田浩一、柘植寛、西田昌史、松井知子、王龍標、話者認識技術の現状と課題、電子情報通信学会音声研究会、2013 年 3 月 1 日、愛知県名古屋
- ④ Kanae Amino, Takashi Osanai, Foreign accent identification using articulation rate of Japanese read speech、14th Australasian International Conference on Speech Science & Technology、2012 年 12 月 6 日、シドニー
- ⑤ 長内隆、法科学分野における音声情報処理、平成 24 年第 5 回情報処理学会東海支部主催講演会(招待講演)、2012 年 11 月 20 日、愛知県名古屋
- ⑥ 長内隆、網野加苗、蒔苗久則、鎌田敏明、母音による性別識別性能の比較、日本法科学技術学会 第 18 回学術集会、2012 年 11 月 16 日、東京都港区
- ⑦ 網野加苗、長内隆、鎌田敏明、蒔苗久則、文発話の調音速度を用いた性別識別、日本法科学技術学会 第 18 回学術集会、2012 年 11 月 16 日、東京都港区
- ⑧ 網野加苗、長内隆、日本語母語話者・非母語話者による読み上げ文章における調音速度、日本音響学会 2012 年春季研究発表会、2012 年 3 月 13 日、神奈川県横浜市
- ⑨ 長内隆、網野加苗、鎌田敏明、蒔苗久則、単音節の子音と母音を用いた話者認識性能の比較、日本法科学技術学会 第 17 回学術集会、2011 年 11 月 18 日、東京都港区
- ⑩ 網野加苗、長内隆、電話番号読み上げ音声を用いた母語・非母語音声識別、日本

- 法科学技術学会 第 17 回学術集会、2011 年 11 月 18 日、東京都港区
- ⑪ 長内隆、網野加苗、鎌田敏明、蒔苗久則、テキスト依存型話者照合における発話内容と話者照合性能、日本音響学会 2011 年秋季研究発表会、2011 年 9 月 21 日、島根県松江市
- ⑫ 網野加苗、長内隆、日本語の電話番号を用いた母語・非母語音声識別、日本音響学会 2011 年秋季研究発表会、2011 年 9 月 22 日、島根県松江市
- ⑬ Kanae Amino, Takashi Osanai、Realisation of the prosodic structure of spoken telephone numbers by native and non-native speakers of Japanese、17th International Congress of Phonetic Sciences、2011 年 8 月 18 日、香港
- ⑭ Kanae Amino, Takashi Osanai、Improvement of likelihood ratio calculation for forensic speaker verification using isolated vowels - Effect of transmission systems-、13th Australasian International Conference on Speech Science & Technology、平成 22 年 12 月 14 日、メルボルン
- ⑮ 長内隆、鎌田敏明、蒔苗久則、網野加苗、テキスト独立型話者照合における発話内容の共通性の評価、日本法科学技術学会 第 16 回学術集会、平成 22 年 11 月 12 日、東京都港区
- ⑯ 網野加苗、長内隆、登録音声サンプル数および内容の話者識別への影響、日本法科学技術学会 第 16 回学術集会、平成 22 年 11 月 12 日、東京都港区
- ⑰ 長内隆、鎌田敏明、蒔苗久則、網野加苗、テキスト独立型話者照合における発話内容の共通性に関する検討、日本音響学会 2010 年秋季研究発表会、平成 22 年 9 月 15 日、大阪府吹田市
- ⑱ 網野加苗、長内隆、話者識別における鼻音・母音の寄与度と母音の鼻音化との関連、日本音響学会 2010 年秋季研究発表会、平成 22 年 9 月 14 日、大阪府吹田市
- ⑲ 長内隆、鎌田敏明、蒔苗久則、テキスト独立型話者照合における話者モデルとその学習に用いるデータ量に関する考察、日本法科学技術学会 第 15 回学術集会、平成 21 年 11 月 12 日、東京都港区

[図書] (計 2 件)

- ① Kanae Amino, Takashi Osanai, Toshiaki Kamada, Hisanori Makinae, Takayuki Arai, Springer Verlag, Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism(Eds.: Amy NEUSTEIN and Hemant A. PATIL) Chap. 1:

Historical and procedural overview of forensic speaker recognition、2011、3-20

- ② Kanae Amino, Takashi Osanai, Toshiaki Kamada, Hisanori Makinae, Takayuki Arai, Springer Verlag, Forensic Speaker Recognition: Law Enforcement and Counter-Terrorism(Eds.: Amy NEUSTEIN and Hemant A. PATIL) Chap.11: Effects of the phonological contents and transmission channels on forensic speaker recognition.、2011、275-308

6. 研究組織

(1) 研究代表者

長内 隆 (OSANAI TAKASHI)
科学警察研究所・法科学第四部・室長
研究者番号：70392264

(2) 研究分担者

なし

(3) 連携研究者

鎌田 敏明 (KAMADA TOSHIAKI)
科学警察研究所・法科学第四部・主任研究官
研究者番号：10356173
蒔苗 久則 (MAKINAE HISANORI)
科学警察研究所・法科学第四部・主任研究官
研究者番号：20415441
網野 加苗 (AMINO KANAЕ)
科学警察研究所・法科学第四部・研究員
研究者番号：70630698

(4) 研究協力者

[海外]

木下 裕子 キャンベラ大学
Phil Rose オーストラリア国立大学
石原 俊一 オーストラリア国立大学

[国内]

・話者認識 SIG メンバー
篠田 浩一 東京工科大学
黒岩 真吾 千葉大学
柘植 覚 大同大学
松井 知子 統計数理研究所
西田 昌史 同志社大学
小川 哲司 早稲田大学
越仲 孝文 日本電気株式会社
王 龍標 長岡技術科学大学 他

[警察関係]

科学捜査研究所音声鑑定担当者