

## Excavation of the Suspect's Identity by Analysis of the Speaker Profiling Case Evidence of Fake Voice Recordings

Tsurayya Yusra Zahirah Zahra<sup>1</sup>; Zhang Wei<sup>2</sup>

<sup>1</sup>Magister of Linguistics, Universitas Gadjah Mada, Indonesia

<sup>2</sup>Foreign Studies, Guangdong University, China

Corresponding author: [tsurayzahra98@gmail.com](mailto:tsurayzahra98@gmail.com)

**Citation:** Zahra, T. Y. Z. & Wei, Z. (2023). Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings. *Notion: Journal of Linguistics, Literature, and Culture* Vol. 5(2), p. 169-178. <https://doi.org/10.12928/notion.v5i2.8038>

### ARTICLE INFO

#### Article History:

Received: 5 Mei 2023

Accepted: 22 September 2023

#### Keywords:

Forensic Linguistics

Speaker Profiling

Fake Voice

Recording

Lesti

Rizky Billar

### ABSTRACT

Technological developments, especially in AI applications, allow someone to commit criminal acts by changing other people's voices. This study implements forensic linguistics by comparing data sources on the audio of Lesti (L) and Rizky Billar (RB) to ensure the originality of the sounds. The Speaker Profiling method with aural-perception analysis is used because there is no valid comparative evidence, so the audio comparison is taken from the L and RB platforms published to the public. This research resulted in two different aspects of each audio: phonetics and acoustics. These two aspects show the two audio sources' differences in pronunciation, emphasis, frequency, and tone. As a result, the voice is identified as a hoax. This research can demonstrate the use of software to help identify individual language profiles; it contributes to the development of applied linguistics, especially as a legal aid tool. Researchers hope that there will be further research from other linguists that will address more benefits of using AI / Corpus technology as a linguistic analysis tool, both with spoken and written data.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

## **I. INTRODUCTION**

The role of speech is not only as a tool to explain an event but also as an investigation tool in the legal realm, as evidence of an investigation or the investigation process itself. Every speech has linguistic strength in an interaction process, regardless of individual social influences, speech power is analyzed through tone, accent, and word choice (Mesthrie, 2011). Speech in the legal process is explained by Barraud in the theory of "énoncé performative" (performative lecture). *Non ce performative est des énoncés qui présentent la singularité d'accomplir ce qu'ils disent du seul fait qu'ils sont dits* (performative speech is speech that has a singularity to the speech said by the speech participants, in order to achieve the purpose of the speech) (Barraud, 2016). The speech contained in an interaction process, especially concerning the use of dialects, is adapted to the social conditions of the community in the speech environment (Hamid & Mboka, 2021). The things that experts say affect the strength of language as an utterance while still considering external factors, namely the environment and the concept of speech function.

External factors become a strong influence on the form, meaning, and function of an utterance, one of the concepts of speech acts is in the realm of law. The existence of language in the legal realm has the power as evidence of a case (Sholihatin, 2017). The legal language / Linguistique Juridique spoken in Barraud's theory (2016) is used by people who are related to legal processes and trials, such as judges, police, suspects, witnesses, to lawyers. These studies are analyzed in a linguistic analysis tool in the legal/forensic realm, which is called linguistic speaker profiling, in forensic linguistics the position of the speaker is the person who is investigated in a case so the term in Indonesian is known as "linguistic profile investigated". (Ahuja & Vyas, 2016).

Speaker profiling is a legal method used by investigators to reveal the suspect's profile by utilizing comparison samples with contemporary samples (Ahuja & Vyas, 2018). The use of this method can be carried out without a comparison sample, provided that there must be the involvement of expert witnesses who are close to the defendant or suspect so that the accuracy of the evidence can be proven valid in the eyes of the law (Watt & Llamas, 2010). The use of speaker profiling followed by the absence of contemporary samples was performed as aural perception (Hollien & Schwartz, 2000). This study applies aural perception to the suspect's video evidence and uploads it on the suspect's YouTube media platform by L (Lesti) and RB (Rizky Billar). In the realm of trial, the victim and the reporter are presented as an expert witness who ensures the validity of the evidence. This study uses linguistic speaker profiling to analyze the linguistic profile of those investigated in the hoax sound recording belonging to L and RB. This can be seen in one of the police investigation tools called "Records of Police Investigations" / « Rapport d'Enquête Policière ».

Le Rapport d'Enquête / Investigation Record is an investigative tool used by investigators to strengthen evidence and facilitate investigations, All words, actions, and matters relating to the investigator will be recorded and easy for further investigation, Recordings of The investigated investigation will also become long-term data from the case in question so that when it is needed, it will be evidence of the credibility of the investigator concerned (Deslauriers-Varin, 2020). On

***Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings***

Investigation Records can be seen the relationship between the victim and the suspect clarifies the investigation. The position of this investigative tool can be a dissertation with evidence of the investigated case, in the form of audio, video, photos, or written evidence relating to the suspect/defendant.

All utterances of the suspect in the form of evidence in the attachment of an Investigation Record were studied using Speaker Profiling. This was done because the researcher saw the number of speeches of the suspect in the case evidence that contained elements of pollution, Sara, and led to acts of fraud. The use of linguistic speaker profiling in a case of Forensic Linguistics was previously discussed by Previous research that also utilized corpus software as a research tool with data sources in the form of sound recordings was Ahuja and Vyas's dialect study, entitled "Forensic speaker profiling: the study of supra-segmental features of Gujarati dialects for text-independent speaker identification". This research shows that one of the functions of using speaker profiling is to explore an individual's identity based on the data source in the form of sound recordings. Ahuja and Vyas' research contributed to forensic research in India by conducting an analysis of the Gujarati dialect with four different regional accents. This study took a sample of 1400 speakers with a ratio of speakers of 50: 50 according to their respective genders. The use of corpus software is needed to carry out phonetic and acoustic element analysis, including the amplitude of vowel sounds which shows the intonation and tone of each Gujarati accent. Identification of dialects using the Speaker Profiling method in Forensic analysis is useful for narrowing down a POI (Person of Interest) or the realm of a search for a defendant or victim.

Ahuja and Vyas' research can show significant differences in the sound of the same words in different accents. One example is the comparison of two accents, where the first accent always ends with a vowel sound, while the second accent always ends with a consonant sound. The word, "Gayaan" means "Singing" and it sounded like "Gayaan" and "Gayuh". The first sound of "Gayan" has an intonation that goes down at the end with one beat, while "gayuh" has a falling intonation from the first syllable and is worth two beats, this can be seen from the description of the amplitude wave of the speech. In addition to detecting intonation, the use of software in dialect analysis has succeeded in identifying differences in speed control and pitch of each Gujarati accent.

In contrast to Ahuja and Vyas' research, this study used aural-perception analysis while sticking to the speaker profiling method. Although both use audio data sources, this research does not have legal audio comparisons, meaning that audio appeals come from sound recordings broadcast publicly without any attachment to the complaint case. This analysis is used to obtain accurate results while still fulfilling the elements of investigative evidence. The comparison video is a 10-minute clip from L and RB's YouTube channel. Comparing L and RB's allegedly hoax voice recordings with the original voices in the comparator audio is the main focus of this study, which aims to establish the linguistic identity of speakers so that voice recordings reported as "hoax voices" can be proven legal and authentic or falsified.

***Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings***

The use of corpus software is expected to clarify acoustic aspects, such as the tone wave in one of the same spoken words in both audio, such as the word "dog", by paying attention to the amplitude wave of the speech when the word is spoken, it will show the emotion and emphasis from each -each speech context in audio one and two. Another thing that will be noticed is the difference in the average frequency and pitch of the sound. Researchers see these two things separately between one speaker and another speaker in each audio.

The average hypothesis is that if the sound has synchronization, then the frequency and pitch will be at relatively the same or close numbers. In addition to using corpus software tools, sound analysis for the creation of linguistic identity is carried out by paying attention to the dialect and idiolect of speakers.

Dialects and idiolects will show the social background and characteristics of the studied language style. Every individual has a style of language that is different from others, therefore forensic linguistic researchers can distinguish voice ownership without knowing the details of an individual's general profile. This research is expected to provide an explanation of the functions and forms of language identification investigated, especially in the data source in the form of sound recordings. In addition, this study also aims to demonstrate the application of technology in language analysis so that the results shown can be more accurate and facilitate the investigation. The more examples of implementing the speaker profiling method, the stronger the position of a forensic linguist will be in the legal process.

## **II. METHODOLOGY**

This research is qualitative with a Forensic Linguistics approach, and the data is in phrases and clauses. The study used 3 stages: (1) data collection, (2) data analysis, and (3) research results presentation. The researcher used the observing and noting method at the data collection stage. Researchers use the observation method to observe usage or behavior in the linguistic domain, accompanied by data recording, to make it easier for researchers to analyze research provisions (Mahsun, 2017).

The researcher uses a language approach method in the realm of law, namely Forensic Linguistics, to analyze the data found in the first stage. This approach is used in linguistic events that have legal indications. At the data analysis stage, the first thing the researcher did was to make a linguistic speaker profiling. After completion, researchers can identify language violations committed by investigators.

The data comes from the YouTube platform, and this is because the L and RB cases have status as public cases, so data transparency has spread across various virtual media. Data collection began with searching for two sound pieces of evidence in the case with comparative data in the form of the actors' original voices (L and RB) via live television recordings. Data selection is based on requirements and a forensic code of ethics, requiring the crime scene to be in the public domain and have comparative evidence.

***Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings***

Analysis of audio recording data using corpus software "Praaline" and "Praat" to see the alignment between the original audio and audio comparison. The use of two corpus software as data analysis tools is based on the following considerations: Praaline can provide more detailed and precise calculations of frequency and pitch compared to Praat, while Praat shows clearer amplitudes than Praaline, features more flexible sound separation in Praat applications compared to Praaline. Finally, the researcher describes his research in the form of a forensic linguistics article.

### **III. RESULT AND DISCUSSION**

The Speaker profiling was used to identify the suspect. In this study, the speaker profiling focuses on linguistic speaker profiling, namely the formation of a suspect's profile based on linguistic elements. Linguistic speaker profiling is used to identify violations or suspect cases based on oral and written language aspects, there are two aspects in it, namely phonetics and speech meaning (Schilling & Marsters, 2015). The meaning of the speech in question is continuous with illocutionary actions in pragmatic studies (Sugiarto & Qurratulaini, 2020), such as agreement speech, apologies, threats, predictions, orders, requests, and others that aim to produce actions from the interlocutor and people related to the speech (Nadar, 2009).

Based on the table of results of speaker profiling of the meaning aspect, from the suspect's speech, five indications of potential language crimes were found that lead to lies, fraud, and defamation. The utterances of the actors fall into the illocutionary act group, this is because the speech contains orders, requests, agreements, and apologies that are addressed to one individual or the public to achieve a certain goal. Analysis of evidence by taking into account the illocutionary aspect is included in the eight basic principles of forensic linguistics: (1) Forensic linguistics bridges two fields of science, namely law, and linguistics, as evidenced by the strong position of linguistics as evidence for legal investigations, (2) linguistic features are used to identify linguistic crimes, (3) linguistic features must be in harmony with legal features, (4) linguistics helps parse linguistic crimes, (5) text and context must support all forensic linguistic investigations, (6) forensic linguistic investigations pay attention to the sequence of crime events, (7) forensic linguistics in two-way collaboration with the investigative team, and (8) identification Forensic linguistics pays attention to aspects of locutions, illocutions, or perlocutions (Gibbons, 2003; Warami, 2014).

#### **Phonetic Aspect**

The analysis carried out on the phonetic aspect consists of an analysis of sound pronunciation and meaning. Sound analysis is based on IPA tables, while meaning analysis uses semiotics and pragmatics based on speech context. This study compares two voice sources belonging to Lesti (L) and Rizky Billar (RB). The first voice recording is a hoax voice recording (allegedly belonging to L and RB), and the second is a voice recording of L and RB originating from a YouTube video on their second channel. The two sound recordings show three of the same words have different pronunciations, including stress and syllable breaks. The three words are "Anjing (dog)", "Lu (you)", and "Engga (no)".

**Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings**

Table 1. Words Identification Lesti (L) and Rizky Billar (RB)

Words	Sound 1 (Hoax)		Sound 2 (Real)	
	L	RB	L	RB
<i>Anjing</i> (dog)	Alphabet "a" pronounce [a] → Vowels central, open.	Alphabet "a" pronounce [ɛ] → Vowels, near-back, near open.	Not Found	The emphasis of the sound is different, found in the letter "i" which sounds [i] → Vowels, central, close.
<i>Lu</i> (you)	Both speakers have a [ɔ] sound to pronounce the letter "u" so it sounds like the word "Lo".		Alphabet "u" pronounce [ʊ] → Vowels, central, close.	Alphabet "u" pronounce [u] → Vowels, back, close.
<i>Engga / gak</i> (no)	Both speakers did word splitting by only saying the last syllable "ga", with the emphasis on the vowel a [ɛ].		Pronunciation without hyphenation with an emphasis on the letter "e", which is pronounced [ə] → Vowels, central, close-mid.	Termination by saying only one last syllable "ga", the letter "a" sound [ɛ] → Vowels, near-back, and near-open.

The results in the table show no significant synchronization of the two audios from the two speakers. The first word "*Anjing* (dog)" has a difference between the first and second audio. The first audio or audio allegedly belonging to L and RB shows an emphasis on the sound of the letter "a", this is different from the second audio, or the original audio which shows an emphasis on the letter "i". In addition, the use of the word "*Anjing*" is not found in L's speech in the second audio. The absence of this word in L's speech means that the word "*Anjing*" is not L's identity.

The second word being compared is the word "*Lu* (you)". This word is a greeting that is intended for the second person pronoun. The word "*Lu*" is mostly used by speakers in the Jabodetabek area (Jakarta, Bogor, Depok, Tangerang, and Bekasi), or Bekasi community groups. Based on the table, there is a correspondence between the two audios, namely the use of the word "*Lu*" as the second person pronoun in L and RB's speech in audios one and two. This shows that the speakers of the two audios have the same social background. However, a significant difference is seen in the pronunciation of the letter "u" which produces two different sounds from each audio. In the first audio, the pronunciation of the letter "u" tends to be open or sounds [ɔ], so it sounds "Lo", while the second audio tends to be closed, so it still sounds "Lu".

Unlike the other two words, the third word shows the most significant difference between the two audios. In the first audio, L's speech and RB's speech show that there is a split in the word "*engga* (no)" to "*ga* (no)." In the first audio, it was suspected that L and RB only said the last word of the word "no," which consists of two letters, while still emphasizing the letter "a." The second audio shows a difference in L's speech. L doesn't cut off the word "*engga* (no)". The speaker

**Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings**

pronounces the word ultimately but emphasizes the letter "e" with a Sundanese accent, so it reads [ə]. While the RB on the second audio has the same sound pattern as the first audio.

**Acoustic Aspect**

The analysis of acoustic aspects is slightly different from other aspects related to linguistics. In the acoustic aspect, there are wave intensity, frequency, and pitch of the audio, therefore corpus software is needed to see accurate results. This study uses two corpus software, namely Praaline and Praat. Both of these software have the same function, which is to help analyze audio files in WAV format. The difference is in the highlighted details, Praaline is more detailed in terms of frequency and pitch. Praaline displays the frequency and pitch according to the desired sound point by adding a color spectrum to clarify the speaker's emotions so that researchers can clearly distinguish the stress in each audio. While Praat emphasizes the "draw system", this makes it easier for researchers to clarify the wave intensity of each audio, Praat can also compare two audios simultaneously.

Table 2. Frequency (F) and Pitch (P) by Praaline

Speaker	Sound 1 (Hoax)		Sound 2 (Real)	
	F	P	F	P
L	818.262 – 990.527 Hz	G#5 – B5	522.180 – 532.947 Hz	C5
RB	1162.79 – 1335.06 Hz	D6 – E6	624.463 – 640.613 Hz	D#5 – E5

The table above shows the frequency range and average pitch of the two audio belonging to L and RB. Frequency and pitch comparison analysis is performed to see the alignment of speech ownership between audio one and two, which helps determine the authenticity of the target audio (first audio). Audio authenticity can be concluded if the frequency and pitch are in a similar or close range of numbers. On the other hand, if the number shown between one audio and another is much different, then the authenticity of the audio can be doubted. Based on the results of the analysis that has been carried out, the frequency and pitch numbers listed are very different from the second audio or the original audio belonging to L and RB, which are used as comparisons.

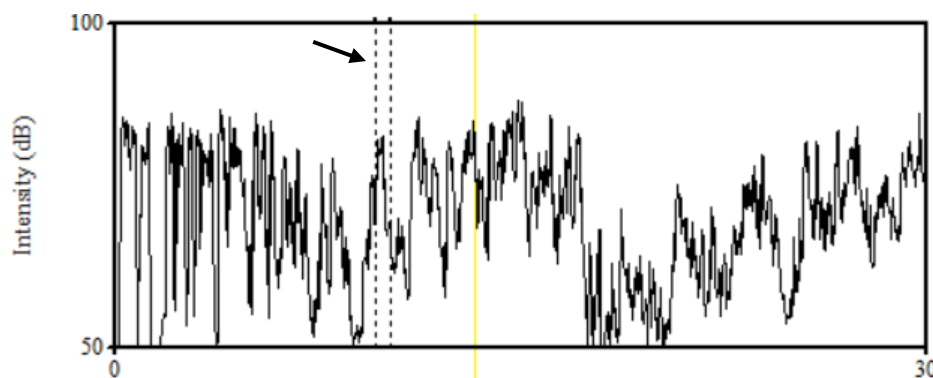


Figure 1. Intensity Wave of RB (Hoax Sound) by PRAAT

**Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings**

The arrows on RB's intensity wave indicate increased speech tone and emphasis on RB's speech. The arrows show the speech when the word "Anjing" is pronounced in the first audio or the audio is suspected to be a hoax belonging to L and RB. Wave height shows an emotional increase when the word is spoken.

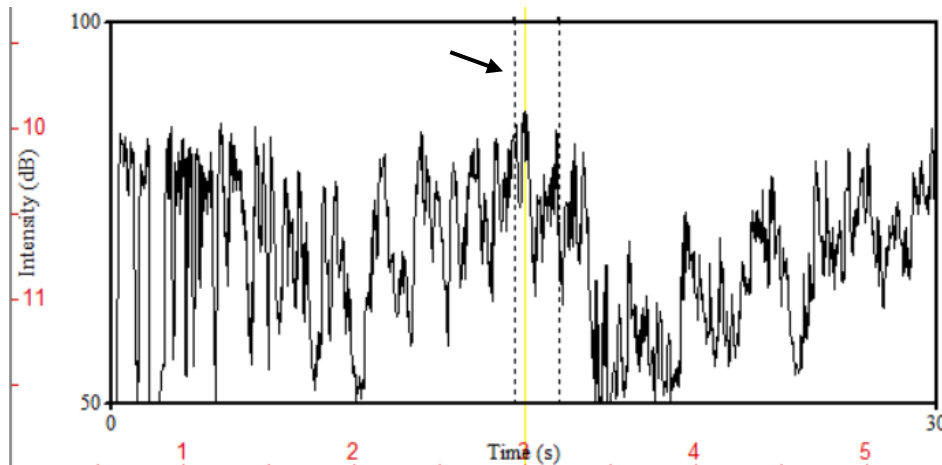


Figure 2. Intensity Wave of L (Hoax Sound) by PRAAT

This is similar to the arrows on L's intensity wave which both show an increase in speech tone and an emphasis on the thickness of the wave seen on the Praat spectrum. The arrow on the L wave also shows the pronunciation of the word "Anjing" in the first audio. These two waves both show the speaker's emotional state with the height and thickness of the intensity wave spectrum belonging to both.

#### **IV. CONCLUSION**

This study shows results that conclude that the audio one or the audio allegedly belonging to L and RB is not the original audio, or not the sound belonging to L and RB. This conclusion was obtained from a phonetic analysis which showed differences in pronunciation and emphasis on the same three words from each audio, and an acoustic analysis which showed significant differences in the frequency and pitch of each audio belonging to L and RB. Researchers hope that there will be further research related to similar cases using other corpus software so that it will add to the scientific contribution to linguistics, especially in the realm of law or forensic linguistics. The contribution of this research belong to both of two studies "law and linguistics".

#### **DECLARATION**

**Author contribution:** The research in this paper is collaborative in the field of forensic linguistics between students (first author), lecturer (second author), and foreign colleagues (third author), all



**Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings**

authors provide their respective perspectives to complement each other's research results in this paper.

**Funding statement:** All funding for this research comes from the personal funds of the first author in accordance with the amount of the author's fee.

**Conflict of interest:** There are no conflict of interest in this paper.

**Additional information:** No additional information is available for this paper.

## REFERENCES

- [1] Ahuja, P., & Vyas, J. M. (2018). Forensic speaker profiling: the study of supra-segmental features of Gujarati dialects for text-independent speaker identification. *Australian Journal of Forensic Sciences*, 50(2), 152-165. <https://doi.org/10.1080/00450618.2016.1237547>
- [2] Barraud, B. (2016). *La linguistique juridique: La recherche juridique*, L'Harmattan. fhal-01367747
- [3] Chaer, A. L. Agustina. (2010). *Sosiolinguistik: Perkenalan Awal: Edisi Revisi*. Jakarta: Rineka Cipta.
- [4] Deslauriers-Varin, N. (2020). Introduction : l'enquête policière et les techniques d'enquête : de la pratique à la recherche. *Criminologie*. 53(2), 5–16. [https://doi.org/10.7202/1074186adresse\\_copiéune\\_erreur](https://doi.org/10.7202/1074186adresse_copiéune_erreur)
- [5] Gibbons, J. (2003). *Forensic linguistics: An introduction to language in the justice system*. John Wiley & Sons.
- [6] Hamid, A., & Mboka, I. (2021). Isolect Language Status of Wakatobi Language in the Sikka Islands (A Dialectological Study). *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, 4(4), 8397-8407. <https://bircu-journal.com/index.php/birci/article/view/2781>
- [7] Hollien, H., & Schwartz, R. (2000). Aural-perceptual speaker identification: problems with noncontemporary samples. *Forensic linguistics*, 7, 199-211. <https://pubmed.ncbi.nlm.nih.gov/11210926/>
- [8] Mahsun. (2017). *Metode Penelitian Bahasa; Edisi ke-3*. Depok: Raja Grafindo Persada.
- [9] Mesthrie, R. (Ed.). (2011). *The Cambridge handbook of sociolinguistics*. Cambridge University Press.
- [10] Nadar, F. X. (2009). *Pragmatik & penelitian pragmatik*. Yogyakarta: Graha Ilmu.
- [11] Satjadibrata, R. (2022). *Kamus Sunda-Indonesia*. Kiblat Buku Utama.

***Excavation of the suspect's identity by analysis of the speaker profiling case evidence of fake voice recordings***

- [12] Schilling, N., & Marsters, A. (2015). Unmasking identity: Speaker profiling for forensic linguistic purposes. *Annual Review of Applied Linguistics*, 35, 195-214. <https://doi.org/10.1017/S0267190514000282>
- [13] Sholihatin, E. (2017). *Pencemaran Nama Baik: Kajian Linguistik Forensik*. Disertasi Pascaunesa.
- [14] Sugiarto, S., & Qurratulaini, R. (2020). Potensi Kriminal Cyber Crime pada Meme: Sebuah Kajian Linguistik Forensik. *Deiksis: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 7(1), 46-57. <https://doi.org/10.33603/deiksis.v7i1.2495>
- [15] Verhaar, J. W. M. (2016). *Asas – Asas Linguistik Umum*; Cetakan ke-9. Yogyakarta: Gadjah Mada University Press.
- [16] Warami, H. (2014). *Legitimasi Kekuasaan dalam UU Otsus Papua: Kajian Wacana Kritis*. Disertasi Doktor. <https://sukarsa.unud.ac.id/repositori/disertasi?nim=1190171005>
- [17] Watt, D., & Llamas, C. (2010). The identification of the individual through speech. *Language and identities*, 76-85. <https://pure.york.ac.uk/portal/en/publications/the-identification-of-the-individual-through-speech>