

In the Matter of

The China Tribunal

Report on Forensic Examinations of Recordings

Dated: 19 February 2020

Author: Professor Peter French

1. SUMMARY

I was asked to provide an analysis and report for the China Tribunal regarding the following matters:

- a) To conduct a comparison of the voice and speech patterns of the recipient of a telephone call made and recorded by an investigator with those of Bai Shuzhong as represented in two video recordings, The purpose of the comparison was to assist the Tribunal in determining whether the recipient of the call was Bai Shuzhong;
- b) To provide a view on whether a portion of the telephone recording in which the investigator may have stated an identity had been edited out/excised;
- c) To determine the number dialled by the investigator from an acoustic analysis of dial tones at the outset of the telephone recording.

I carried out the work on a *pro bono publico* basis with the assistance of colleagues named in the body of the report. The methods of examination I employed are widely used and accepted within the forensic speech and audio community.

My conclusions are:

Speaker Comparison: *The voice in the questioned telephone call is very similar to and consistent with that of Bai Shuzhong. I found no differences that would indicate that the voice in the call is not his. However, I am unable to estimate how many other adult male speakers of the same dialect of Mandarin Chinese would also share the voice features found in the call recording.*

Editing out of Material: *I am unable to provide a view on this.*

Number Dialled: *00861051817608.*

2. PERSONNEL

I am Professor of Forensic Speech Science at the University of York, Visiting Professor in the same subject at the University of Huddersfield and at York St John University, Executive Board member at the Centre for Forensic Phonetics and Acoustics at the University of Zurich and Chairman of JP French Associates Forensic Speech and Acoustics Laboratory. Over the past 35 years I have been involved in more than 5000 legal cases in the UK and in other countries across the world in which expert analysis of speech and audio recordings was required. I have been enlisted for matters heard before courts of all levels, including International Commission of Enquiry and International Criminal Tribunal for the Former Yugoslavia

I was assisted in the present case by [name] who is employed as a Research Assistant and Graduate Teaching Assistant in [institution], where they are completing a PhD in Linguistics. Although their first language is Cantonese, they have a very high level of proficiency in Mandarin (the language used in the recordings in this case), having learned and spoken it from the age of 4 years.

Details of other staff who assisted are given in footnotes to Section 3 below.

3. INSTRUCTIONS

I was initially contacted by email on 7th February 2020 and asked to provide expert input to the China Tribunal by comparing the voice of the recipient of a telephone call, claimed to be Bai Shuzhong, with the known voice of Bai Shuzhong heard in two video recordings. The purpose of the comparison was to assist the Tribunal in deciding whether the recipient of the call was Bai Shuzhong.

I was subsequently informed that the telephone call, made in 2014, could be found at the following internet link:

http://www.zhuichaguoji.org/sites/default/files/investigation_report/2014/09/44818_baishuzhong10sep-2-final-mp3.mp3

and that the 2 known video files of Bai Shuzhong, recorded respectively in 2019 and 2015, could be found at these links:

http://jk.ccntv.cn/html/2019/jkft_0320/1710.html

<https://v.qq.com/x/page/z0168bkname.html>

In addition to the forensic speaker comparison work, I was asked if I could provide a view on whether a section of the telephone recording in which the investigator initiating the call had given his name had been edited out, and whether I could determine from the recording the number dialled by the investigator.

It was agreed that, owing to the very short time between my date of instruction and the deadline for a report (c. 10 days), plus intervening commitments, a very detailed report with full quality assurance measures could not be provided.

4. PREPARATORY WORK

The telephone recording and the 2019 video recording were downloaded from the internet. The 2015 video recording could not be downloaded, and was therefore re-recorded via a digital link. The telephone recording and the audio from the video recordings was converted to, and stored in, standard audio data (.wav) format. This work was undertaken by my colleague, Dr Philip Harrison¹.

For the speaker comparison tests, I edited each of the sound files in order to remove the voices of speakers other than the recipient of the telephone call and Bai Shuzhong in the video recordings. The video recording of 2015, although being nearer in time to the questioned telephone call, contained speech that was very formal in style and appeared to be following a prepared script. There was a large mismatch in speaking style between this recording and that found in the questioned call. The 2015 recording was also marred by a certain degree of reverberation. The video recording of 2019 was therefore selected as the

¹ Forensic Consultant and Director of JP French Associates, Lecturer in forensic speech science at the University of York.

main reference sample in this case, although the earlier recording was also taken into account in arriving at my conclusion.

Telephone transmitted speech is of limited frequency range. The speech from the 2019 video was therefore digitally filtered in order to emulate the signal frequency characteristics of the telephone material, thus allowing a more direct comparison². Both the filtered and non-filtered versions were used during my analysis.

5. EXAMINATIONS

Adequacy of material

Although subject to the usual frequency bandwidth limitation of telephone transmitted speech, the call recording was found to be of clear quality for material of this kind.

As stated above, the 2015 video recording was somewhat marred by reverberation, but the 2019 video recording was quite free of factors that might inhibit analysis.

I have taken the limitations of the samples into account in arriving at my conclusions.

Analysis methods

With the assistance of [name], I analysed the recordings using auditory-phonetic (analytic listening) and acoustic-phonetic (computer-based measurement of physical aspects of the speech signal) methods.

Consonant pronunciations were examined and compared, as were vowels using speech spectrography and procedures that allow one to log and represent their first and second constituent resonances (formants) on overlaid scatter plots.

The third vowel formant, a good marker of individual identity, was averaged and compared across the samples from the mixture of all vowels analysed.

² The filtering was undertaken by my colleague, Dr Jessica Wormald, Forensic Consultant at JP French Associates.

Voice pitch was examined and compared as fundamental frequency – fundamental frequency is the acoustic correlate of perceptual pitch and refers to the rate of vocal cord vibration estimated automatically by computer programme and averaged over series of utterances.

Voice quality (timbre, or the individual ‘colouring’ of the voice) was analysed and compared using a formalised system of categories, a version of the Edinburgh Vocal Profile Analysis scheme adapted for forensic purposes.

A list of the equipment (hardware and software) used in the examinations is appended to this report.

6. RESULTS

I found a range of similarities between the voice and speech patterns in the 2019 video recording and those in the questioned call. There were no forensically significant differences, i.e., differences which could not be explained by the different recording characteristics, and natural variation in speech from one individual. Findings included:

Voice quality. The questioned and reference recording voice qualities were found to be very similar, both at a componential and a holistic level. This was especially apparent when comparing the questioned telephone call and the filtered version of the 2019 video material.

Voice quality was characterised by the following combination of features:

- harshness
- low level of breathiness
- creakiness in the lower part of the pitch range
- laryngeal tension
- slightly raised larynx
- nasality
- fronted tongue body
- intermittent retroflexion

Voice pitch. The mean fundamental frequency value (f_0) obtained for 71 seconds of speech extracted from the telephone call recording was 119 Hz (SD = 32 Hz) and that from 347

seconds of speech from The China Tribunal was 115 Hz (SD = 33 Hz). The difference between the two samples is small and well within the range one would expect to find when comparing speech from the same person talking in the two different situations.

Segmental features. The questioned speaker in the telephone recording and Bai Shuzhong in The China Tribunal speak with a Northern variety of Mandarin Chinese. They have a range of similarities in respect of pronunciations, including, *inter alia*, retroflex fricative realisations of /ʃ/ and variably palato-alveolar ~ retroflex realisations of the affricate /tʃ/, and rhoticisation of coda position vocalic segments.

Multiple tokens of six Mandarin monophthongal vowels were identified and compared acoustically across the reference and questioned samples. Similarities were observed in respect of the overall distributions within the F1 - F2 'vowel space' and within vowel categories. Average F3 values were also similar across the samples and differences were within the range that one might expect from the same person speaking on different occasions and in different circumstances.

Tones. Mandarin being a lexically contrastive tone language, the use of tones across the recordings was also considered. While no idiosyncratic uses were found, neither were there any significant differences across the samples.

In respect of the parameters considered overall, no forensically significant differences were found.

7. SPEAKER COMPARISON CONCLUSIONS

In interpreting findings and drawing conclusions in speaker comparison cases, one normally considers the degree of *similarity* between the reference and questioned recordings. Similarities support the view that the speakers are the same, whereas differences support the view that they are different.

One also considers the *typicality* or *distinctiveness* of the features common to the questioned and reference recordings, as this has a bearing on how many speakers of a comparable variety of the language might be expected to share these features. More distinctive or unusual

similarities mean the evidence is stronger, as fewer other people would also share those features.

In the present case, it has been possible to assess the degree of similarity across the telephone recording and the known recording of Bai Shuzhong. They are very similar. However, neither [name] nor I has any population data relating to, nor sufficient previous forensic casework experience of, Mandarin for making a judgement of typicality/distinctiveness. That being the case, the most I can say is that *the voice in the questioned telephone call is very similar to and consistent with that of Bai Shuzhong in respect of the parameters examined; I have found no differences that would indicate that the voice in the call is not his. However, I am unable to estimate how many other adult male speakers of the same dialect of Mandarin would also share the voice features found in the call recording.*

8. FORENSIC SPEAKER COMPARISON EVIDENCE: STANDARD CAVEATS

(a) Forensic speaker comparison tests do not ordinarily provide categorical statements of identification (or ‘matches’). This is because it is recognised that there could be people in the population who are indistinguishable in respect of voice and speech patterns, people’s voices vary in different situations, and voices can also be affected by different recording conditions.

(b) While the examinations are based on established scientific methods of analysis, the conclusion is my expert opinion, i.e., an interpretation of the findings based on my specialist education, training and experience of analysing speech (rather than by statistical testing using a database, for example).

9. EDITING

I am unable to determine whether a segment regarding the identity of the investigator making the questioned call has been edited out of the telephone recording. Authenticity examinations of that kind normally require access to the original recording, the device on which the recording was made and, potentially, access to electrical frequency network data relating to the time when it is claimed the recording was made from the area in question. Such examinations would normally be conducted by an acoustic engineer. At JP French Associates this would be done by my colleague, Dr Philip Harrison.

10. NUMBER DIALLED

At the beginning of the call recording there are a number of dialling tones - technically, Dial Tone Multi Frequencies or DTMFs. I was given no information about what the number was claimed to be but have analysed the DTMFs acoustically and found the number dialled by the investigator to be **00861051817608**. This was independently checked and confirmed by Dr Harrison.



Professor Peter French

19th February 2020

APPENDIX – EQUIPMENT USED

Hardware

HP 400 G3 Windows-based PC
Focusrite Scarlett 2i2 USB audio interface
M-Audio M-Track Hub audio interface
Samson S-phone headphone amplifier
Beyerdynamic DT250 headphones

Software

Sony Sound Forge (version 9.0e)
Praat (version 6.0.22) with in-house modifications for forensic purposes