

FORENSIC VOICE COMPARISON USING CHINESE /iau/

Cuiling Zhang^{a,b}, Geoffrey Stewart Morrison^b & Tharmarajah Thiruvanan^b

^aDepartment of Forensic Science & Technology, China Criminal Police University,
Shenyang, China;

^bForensic Voice Comparison Laboratory, School of Electrical Engineering & Telecommunications,
University of New South Wales, Sydney, Australia

cuilingzhang@yahoo.com.cn; geoff-morrison@forensic-voice-comparison.net;
thiru@ee.unsw.edu.au

ABSTRACT

An acoustic-phonetic forensic-voice-comparison system extracted information from the formant trajectories of tokens of Standard Chinese /iau/. When this information was added to a generic automatic forensic-voice-comparison system, which did not itself exploit acoustic-phonetic information, there was a substantial improvement in system validity but a decline in system reliability.

Keywords: forensic voice comparison, acoustic-phonetic, automatic

1. INTRODUCTION

A number of studies, e.g. [5, 10, 14, 19] conducted within the new paradigm for forensic-comparison science [9, 12] have explored the effectiveness of acoustic-phonetic forensic voice comparison based on the coefficient values of parametric curves fitted to the formant trajectories of diphthongs. The earlier studies used controlled (e.g., read) speech, and did not investigate whether this technique lead to improvements over an automatic system. It is important to ascertain whether the expense of the human labor involved in acoustic-phonetic procedures is justified by substantial improvement in performance over a cheaper automatic system.

The present paper explores the effectiveness of the formant trajectory technique applied to tokens of the Standard Chinese triphthong /iau/ on tone 1 (high level), occurring in the word — *yao* “one”. Triphthongs being more complex than diphthongs, it may be possible to extract more useful information from a triphthong than a diphthong. The tokens are extracted from a database of spontaneous speech, being more forensically realistic in this respect than the controlled speech of earlier studies. The effectiveness of the formant-trajectory technique is assessed as the improvement in performance obtained when the

latter is added to a generic automatic forensic-voice-comparison system. Improvement in performance is measured using the log-likelihood-ratio cost (C_{lr}) as a metric of validity [4, 7] and a parametric estimate of the 95% credible interval (CI) for the likelihood ratios (LRs) as a metric of reliability [15, 19, 20].

2. METHODOLOGY

2.1. Data

The data were extracted from a database [23] of voice recordings of female speakers of Standard Chinese (a.k.a. Mandarin and Putonghua). See [17] for details of the data collection protocol. The data consisted of 2 recordings of each of 60 speakers. The speakers were all first-language speakers of Standard Chinese from northeastern China, and were aged from 23 to 45 (with most being between 24 and 26). The recordings used were from an information exchange task conducted over the telephone: Each of a pair of speakers received a “badly transmitted fax” including some illegible information, and had to ask the other speaker to provide them with the missing information. The original recordings were approximately 10 minutes long, with the second recording of each speaker recorded 2-3 weeks after the first. Recordings were high quality, recorded at 44.1 kHz 16 bit using flat-frequency response lapel microphones. The present paper should be considered a preliminary to testing more forensically realistic conditions including transmission-channel and speaking-style mismatches.

Data from the first 20 speakers (01-04, 09-20, 22, 25, 26, 28) were used as background data, data from the next 20 speakers (29-48) were used as development data, and data from the last 20 speakers (49-68) were used as test data.