

Automatic Telephone Voice Analysis and Therapy

Ladan Baghai-Ravary and Steve W. Beet
Advanced Speech Processing Team, Aculab plc

{ ladan.ravary, steve.beet } @ aculab.com



Aim

To analyse a speaker's voice over the telephone, for monitoring of patients with voice disorders and to provide Speech and Language Therapists with an automated facility for practice sessions with their clients. The analysed features can be selected from a large variety of "VoiScan" parameters [1].

Parameters

The VoiScan parameters include many traditional voice and speech analysis parameters, as well as some novel ones. The measures include:

- Pitch
- Jitter
- Shimmer
- Breathiness
- Creakiness
- Voice break rate
- Formant parameters
- Hirschman uncertainty
- Spectro-temporal entropy
- Spectro-temporal dynamics

Calibration

The VoiScan parameters for an individual can be compared with population statistics which have been calculated from 20,000 telephone calls in 12 countries.

Validation

The sensitivity of the VoiScan parameters to telephone distortions has been investigated by comparing results from high-quality recordings with those from simultaneous recordings made over a normal telephone landline [2].

The VoiScan parameters were significantly more robust than the standard Praat [3] equivalents. The VoiScan pitch, shimmer and noise-to-harmonic ratio (NHR) were especially robust.

The Pearson R^2 values between VoiScan and Praat parameters of high quality recordings, were all in excess of 0.9.

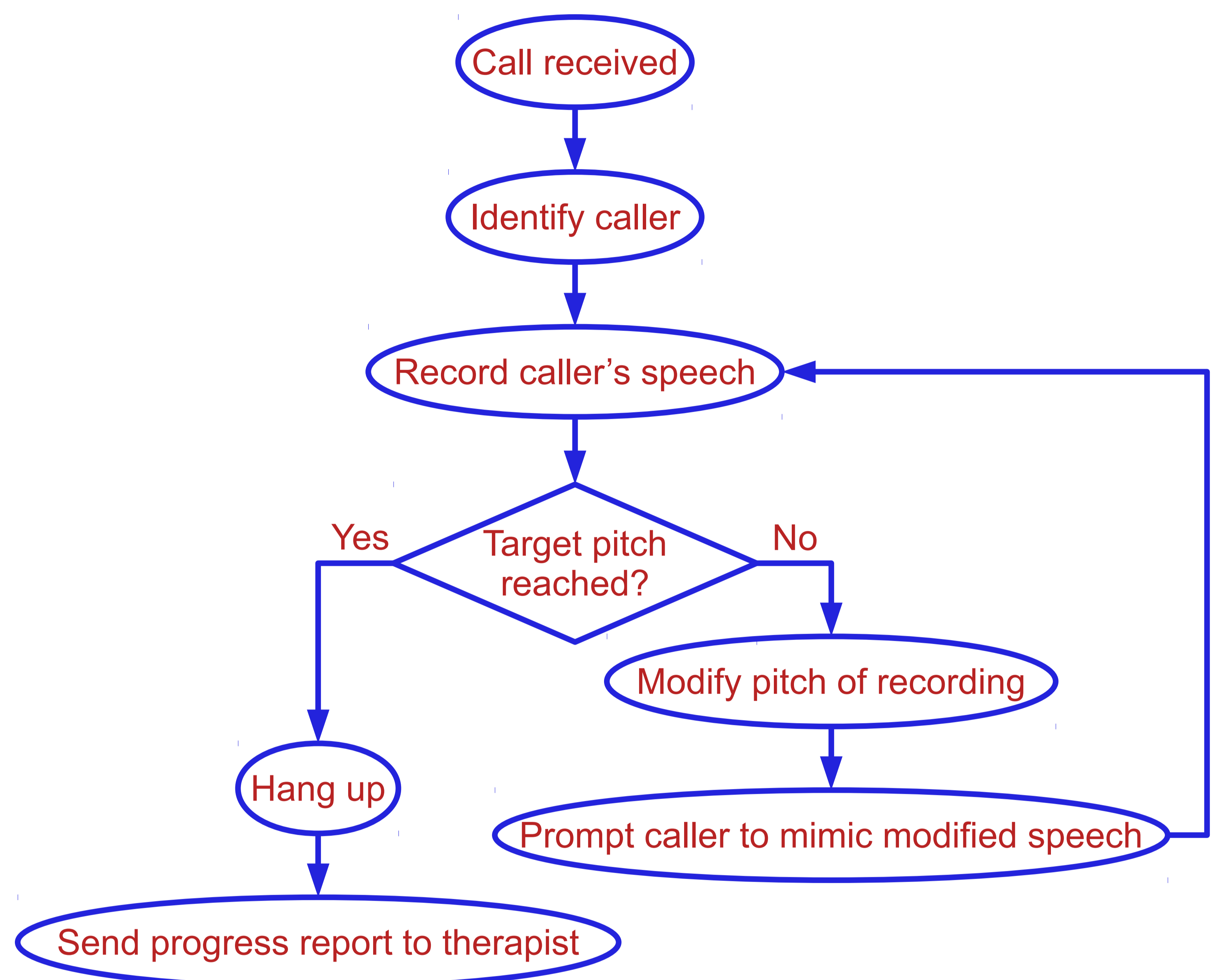
Conclusions

- The system can be readily configured to allow a therapist to monitor other aspects of a client's speech, including numerical indices corresponding to breathiness and creakiness.
- The VoiScan parameters are sufficiently robust for everyday telephone use, and can be derived from sustained phonations or even humming, as well as from natural or read speech.
- Clients can practice frequently and regularly, and over an extended period of time, without the need for intervention by the therapist.
- The system has been implemented using cloud telephony, making it cost-effective and scalable.

References

1. "Aculab Cloud voice and speech analysis system parameters and algorithms", Aculab plc, https://www.aculab.com/products/aculab-cloud_telephony_platform/clinical-speech-analysis-system/, last accessed 5 September 2017.
2. "VoiScan: Telephone Voice Analysis for Health and Biometric Applications", L. Baghai-Ravary and S. W. Beet, to appear in "Speech and Computer", Springer, 2017.
3. "Praat: doing phonetics by computer", <http://www.praat.org/>, P. Boersma and D. Weenink, last accessed 5 September 2017.

Pilot Study: Pitch Control



Technologies

- Interactive Voice Response (IVR)
- Voice and Speech Analysis
- Voice Recording and Playback
- Speech Recognition (ASR)
- Speech Synthesis (TTS)
- Voice Manipulation
- Key-press (DTMF) Detection
- Tone Generation
- Text Messaging (SMS)
- Web Services / Email

Results

- Typically, even on a client's first call, they will take only a couple of minutes to change their pitch by 4 semitones or more.
- The extent of the change achievable, and the speed with which it is achieved, both increase rapidly with practice.
- The target pitch and the required precision and stability of the pitch can be customised for individual clients.