Improving presentation intonation with feedback on pitch variation

Rebecca Hincks
Unit for Language and Communication
KTH
hincks@kth.se

Jens Edlund
Centre for Speech Technology
KTH
edlund@speech.kth.se

The use of speech analysis in teaching second language discourse intonation has traditionally relied upon the visual display of pitch contours over short segments of speech. A number of researchers have pointed out that this method has considerable limitations. We report our test of the idea that valuable information about learner intonation can be gained by circumventing the contour visualization process altogether and processing the distribution of the pitch data only. We have developed a system that gives online, instantaneous and transient feedback on the standard deviation of fundamental frequency as measured in the perceptually relevant semitone scale.

Seven native speakers of Mandarin have practiced oral presentations with the feedback and been compared on the basis of change in pitch variation levels with a control group of seven speakers. Both groups increased their pitch variation with training, and the effect lasted after the training had ended. The test group showed a significantly higher increase than the control group, indicating that the feedback is effective. There was however a great variation between individuals.

Our technique attempts to address the problems encountered by those second language users, such as Asian teaching assistants at American universities, who have difficulty making full expressive use of their pitch ranges as they speak. It is conceived as a component in a computer environment for practicing oral presentations (Hincks, 2005). Like the majority of CALL systems, a presentation practice system would apply behaviorist theories of learning by providing environments for skills practice where learners are rewarded for meeting certain targets. Unlike most CALL systems, however, here the student input will be freely generated speech with an authentic communicative intent. Enabling communication with a computer is no simple matter, yet much research points to the supremacy of communicative techniques when it comes to teaching a second language. Having the computer respond to the prosody of presentation speech rather than its lexical content is one way of having it react to the communicative intent of the speaker. In such a system, the target levels for prosodic variation would be flexible, allowing for instructional scaffolding in response to the initial skills of the learner. By providing an environment for rehearsing a presentation, the system would encourage the use of self-assessment by allowing learners to record themselves as they practice. Many learners are bewildered by advice such as: 'use more variation in your speaking style;' such a system would allow them to test different styles on their own. Finally, like many applications of information and communication technologies in learning situations, the application would stimulate lifelong learning, by being available to users outside traditional classroom settings.

The system used in the present experiments consists of a base system allowing students to listen to teacher recordings (targets), read transcripts of these recordings, and to make their own recordings of their attempts to mimic the targets. Students may also make recordings of free readings. Furthermore, students can browse through targets, make new recordings and listen to their latest recording. The interface keeps track of the students' actions, and some of this information, such as the number of times a student has attempted a target, is continuously presented to the student.

The base system is extended with a component providing online, instantaneous and transient feedback visualizing the degree of pitch variation the student currently produces. The feedback is presented in a meter that is reminiscent of the amplitude bars used in equalizers: the current amount of variation is indicated by the

number of bars that are lit up in a stack of bars, and the highest variation over the past two seconds is indicated by a lingering top bar. The meter has a short, constant latency of 100ms

The meter is fed data from an online analysis of the recorded speech signal. The analysis used in these experiments is based on /nailon/ online prosodic analysis software and the Snack sound toolkit. As the student speaks, a fundamental frequency estimation is continuously extracted using an incremental version of getF0/RAPT (Talkin, 1995).The estimation frequency is transformed from Hertz to logarithmic semitones. There are several reasons for this transformation: semitones are perceptually relevant, which moves us from fundamental frequency (an acoustic measure) to pitch (a perceptual measure; semitones are perceptually equidistant, so that a rise of one semitone from five to six is perceptually the same as a rise from 13 to 14). This gives us a kind of perceptual speaker normalization. Fundamental frequency distributions over a single speaker also fit a normal distribution more closely if measured in semitones than in Hertz (Edlund & Heldner, 2007), making the following steps more reliable.

The next step is a continuous and incremental calculation of the standard deviation of the student's pitch over the last 10 seconds. The result is a measure of the student's recent pitch variation. This value is normalized against a base value (in the present experiments, this is the pitch variation the student produced in the initial session). Finally, the meter utilizes a dampening function, making it impossible for students to max the meter out – the more bars that are lit up, the more variation that is needed to light another one. The pitch meter shows yellow bars when the pitch variation is low or similar relative to the student's initial reading, and green bars when it is higher

While a system of this nature cannot tell a learner whether he or she has made pitch movement that is specifically appropriate or native-like, it should stimulate the use of more pitch variation in speakers who underuse the potential of their voices to create focus and contrast in their instructional discourse. It can be seen as a first step toward more native-like intonation, and furthermore to becoming a better public speaker. In analogy with other learning activities, we could say that a system like ours aims to teach students to swing the club without necessarily hitting the ball perfectly the first time. Most importantly, because the system gives feedback on the production of free speech as well as specific utterances, it stimulates and provides an environment for the practice of authentic communication such as the oral presentation.

References

Edlund, J., & Heldner, M. (2007). *Underpinning /nailon/ - automatic estimation of pitch range and speaker relative pitch*. In C. Müller (Ed.), Speaker Classification I: Fundamentals, Features, and Methods: Springer.

Hincks, R. (2005), *Computer Support for Learners of Spoken English*. Doctoral Thesis, TMH, KTH.

Talkin, D. (1995). *A robust algorithm for pitch tracking (RAPT)*. In W. B. Klejin, & Paliwal, K. K (Ed.), Speech Coding and Synthesis (pp. 495-518).