

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

System xxx (2005) xxx–xxx

SYSTEM

[www.elsevier.com/locate/system](http://www.elsevier.com/locate/system)

2 Measures and perceptions of liveliness in student  
3 oral presentation speech: A proposal for  
4 an automatic feedback mechanism

5 Rebecca Hincks

6 *Department of Speech, Music and Hearing, The Royal Institute of Technology (KTH), SE100*  
7 *44 Stockholm, Sweden*

Received 19 January 2005; received in revised form 18 April 2005; accepted 24 April 2005

---

10 **Abstract**

11 This paper analyzes prosodic variables in a corpus of eighteen oral presentations made by stu-  
12 dents of Technical English, all of whom were native speakers of Swedish. The focus is on the extent  
13 to which speakers were able to use their voices in a lively manner, and the hypothesis tested is that  
14 speakers who had high pitch variation as they spoke would be perceived as livelier speakers. A metric  
15 (termed PVQ), derived from the standard deviation in fundamental frequency, is proposed as a mea-  
16 sure of pitch variation. Composite listener ratings of liveliness for nine 10-s samples of speech per  
17 speaker correlate strongly ( $r = .83$ ,  $n = 18$ ,  $p < .01$ ) with the PVQ metric. Liveliness ratings for indi-  
18 vidual 10-s samples of speech show moderate but significant ( $n = 81$ ,  $p < .01$ ) correlations:  $r = .70$  for  
19 males and  $r = .64$  for females. The paper also investigates rate of speech and fluency variables in this  
20 corpus of L2 English. An application for this research is in presentation skills training, where com-  
21 puter feedback could be provided for speaking rate and the extent to which speakers have been able  
22 to use their voices in an engaging manner.

23 © 2005 Elsevier Ltd. All rights reserved.

24 *Keywords:* Prosody; Intonation; Pitch; Speech rate; Oral presentation skills; CALL; Speech analysis; English for  
specific purposes; Fluency; Learner corpora

26

---

---

*E-mail address:* [hincks@speech.kth.se](mailto:hincks@speech.kth.se)

## 27 1. Introduction

28 Any number of popular manuals on public speaking (e.g., Lamerton, 2001; Grandstaff,  
29 2004) advise speaking with a lively voice that varies in intonation. The word 'lively' means,  
30 if one consults Merriam-Webster (online), 'briskly alert and energetic' and 'imparting spirit  
31 and vivacity.' Collins COBUILD (1995) associates liveliness primarily with enthusiasm.  
32 According to the manuals, a lively voice is achieved by consciously modifying the three pro-  
33 sodic dimensions of loudness, pitch and tempo. Intonational modification helps the audience  
34 understand the content of the message. By pausing before moving to a new point, for exam-  
35 ple, and then raising pitch as one starts to speak, a speaker helps listeners orient themselves in  
36 the flow of information. An important side effect of helping listeners in this way is the main-  
37 tenance of listener focus on the message, so that their attention does not wander. Lively  
38 speakers should also avoid following one intonational pattern utterance after utterance,  
39 and include a visual dimension, with the contribution of facial and body gestures. Using one's  
40 voice well is not the absolutely most critical aspect of making a good presentation; obviously  
41 it is also important that the content is well-structured, appropriate to the audience, and  
42 clearly explained. Yet if the speaker does not use his or her voice in a way that facilitates ac-  
43 cess to the content, much of the message can be lost. It is surprising that the area has attracted  
44 little academic interest; a plausible reason for this is that it is only recently that technological  
45 development has allowed smooth processing of large amounts of recorded speech.

46 The quantity of self-help books on presentation skills on the market today<sup>1</sup> is testimony  
47 both to the demands put on oral communication in contemporary workplaces and to peo-  
48 ple's lack of preparation to meet these demands. Not every graduate has had the oppor-  
49 tunity to take a course in speaking skills. Speakers who turn to self-help manuals are told  
50 to practice on their own or with a friend. In his chapter entitled "Improving academic and  
51 medical presentations" one expert (Grandstaff, 2004) gives this advice:

"If you are not sure whether you spoke in a monotone, record yourself and listen for how much variety you use in your voice as well as whether you are speaking faster or slower than speakers you enjoy. Ask a friend or colleague to listen to the tape and give suggestions about how you can add interest and variety to your voice. Practice by varying your pitch, pace and volume. Make the variety fit what you are saying. Emphasize key words. Pause to add impact and to allow time for people to take in what you have said. Increase or lower your volume slightly to draw attention to key words or phrases." (p. 237)

60 This paper suggests that a computer could fill the role of friend or colleague and give  
61 automatic, objective and valuable feedback on speaker prosody. Speech analysis software,  
62 which has been used for the past 25 years to help second language learners visualize the ways  
63 in which their intonation deviates from a target model (Bot and Mailfert, 1982; Molholt,  
64 1988; Anderson-Hsieh, 1992; Öster, 1998; Hardison, 2004), can also be used to gather  
65 raw data about pitch variation, speaking rate and pausing. The question underlying the re-  
66 search reported here is whether such data can be used to characterize speaker liveliness. If it  
67 could, then a potential new application for speech analysis software would be as a feedback

<sup>1</sup> In November 2004, Stockholm's largest bookstore had three times as many shelves full of books on advice for public speaking as on writing.

68 or evaluative mechanism for public speaking. The paper focuses exclusively on the variables  
69 that can be collected and processed automatically and online; that is, without reference to  
70 propositional content. In other words, I will not address, at this point in technological devel-  
71 opment, Grandstaff's advice to 'make the variety fit what you are saying.' However, I feel  
72 that many speakers would still benefit by the kind of feedback I am proposing.

73 Public speaking difficulties are magnified for second language users, who are operating un-  
74 der a heavy cognitive load of planning lexical content and its articulation at the same time as  
75 they may lack confidence and familiarity with the potentialities of spoken academic English.  
76 A number of recent works (Ventola et al., 2002; Rowley-Jolivet and Carter-Thomas, 2005)  
77 have pointed to the lexical, syntactical and pragmatic difficulties faced by non-native speakers  
78 who are presenting or teaching in English. An increasing number of studies have addressed  
79 the important issue of non-native prosody in instructional speech (Hahn, 2004; Levis and  
80 Pickering, 2004; Pickering, 2004). Pickering (2004) compared the way native and non-native  
81 teaching assistants used intonational paragraphing (Brazil, 1997) in the presentation of lab-  
82 oratory instructions. The non-native speakers showed "a considerably weaker control of  
83 intonational structure and a disturbance in prosodic composition that materially affects  
84 the comprehensibility of the discourse for native speaker hearers" (Pickering, 2004, p. 19).  
85 One of the contributing problems was an overall narrower pitch range, which made the iden-  
86 tification of prosodic units difficult. International teaching assistants are responsible for a  
87 large amount of undergraduate instruction at many North American universities, and their  
88 communication difficulties are a serious problem. Yet, as Pickering notes,

"little may be done to...address areas of linguistic competence such as pitch range or  
pause structure, as they are often perceived to be less crucial for functional compe-  
tence than lexical or syntactic marking strategies... However, prosodic cues contrib-  
ute independently to the structure of the discourse, and they cannot be circumvented  
without a reduction in comprehensibility" (Pickering, 2004, p. 39).

94 Ideally, an instructor or speaker who faces such challenges would be given individual-  
95 ized expert instruction in how to improve his or her prosody in relation to the content of  
96 the message. Since this kind of coaching is probably not realistic in most university set-  
97 tings, an alternative would be participation in a course in speaking and presentation skills.  
98 Pickering suggests that such courses could benefit from the introduction of exercises de-  
99 signed for theater training as a potential remedy for the problem of restricted pitch range  
100 (ibid p. 39), and Levis and Pickering (2004) suggest using speech analysis for visualization  
101 of pitch contours at the discourse level.

102 As Levis and Pickering point out, speech visualization has too often been used to prac-  
103 tice intonation only at the level of phrases or short utterances. When a learner does this, he  
104 or she is imitating the pitch contour of a specific speaker. This entails adopting that speak-  
105 er's attitude and dialect. Such imitation might be helpful to beginning learners, but can be  
106 of limited benefit in the long run. Jenkins (2000) claims that speakers of international Eng-  
107 lish do not need the kind of intonational training designed to make them sound like na-  
108 tives, though speakers must master the placement of focus<sup>2</sup> in an utterance. This finding

---

<sup>2</sup> Both Jenkins and Hahn use the terms 'nuclear stress' and 'primary stress' for what other scholars call 'focus'. Following Wennerstrom (2001) and others I reserve the terms 'stressed' and 'unstressed' to describe the relationships of syllables to each other at the lexical level, and 'focus' to describe relationships of syllables at the utterance level.

109 was corroborated by Hahn (2004). The system described here would not be able to deter-  
110 mine whether the placement of focus was correct or not, but it would help point to whether  
111 the speaker was expressing focus at all. In Hahn's research, subjects were presented with  
112 non-native instructional speech in three conditions: delivered with normal sentence focus,  
113 abnormal sentence focus, and without focus (monotone). Students comprehended and re-  
114 called more information from the correctly focused delivery, and in evaluative comments,  
115 were most critical of the focus-less delivery. Interestingly, 30% of the students who listened  
116 to the focus-less delivery thought that the speaker spoke too fast, though speaking rate  
117 and pausing were tightly controlled and no subjects who had heard deliveries with focus  
118 commented on speaking rate.

119 There is a further, practical reason that speech analysis has been used to visualize only  
120 short utterances: if the pitch contour of a very long utterance is shown on one computer  
121 screen, many important details and movements are lost by being compressed. Therefore,  
122 this paper advocates looking at the distribution of the pitch data only, without visualiza-  
123 tion. The proposed feedback mechanism processes large amounts of pitch data in terms of  
124 the standard deviation of the fundamental frequency in order to detect the degree to which  
125 the speaker was varying his or her pitch over long stretches of discourse.

126 Ultimately, one can envision a feedback mechanism for presentation speech that incor-  
127 porates speaker-dependent speech recognition to recognize, and process at some level, the  
128 linguistic content of the presentation or lecture. Using natural language processing, the  
129 instant transcript could be textually analyzed for features that have been deemed appropri-  
130 ate for the speaking genre in question. The recent attention paid to spoken academic Eng-  
131 lish has helped our understanding of the lexical and syntactic properties of successful  
132 monologue (Camiciottoli, 2003; Simpson et al., 2003; Morell, 2004; Rowley-Jolivet and  
133 Carter-Thomas, 2005). Like grammar checkers that can be set to flag certain stylistic fea-  
134 tures of written texts, this feedback mechanism could check for the presence of personal  
135 pronouns (a positive feature in monologue (Morell, 2004)) or passive constructions (a neg-  
136 ative feature, indicating difficulties in adapting the information structure of a written text to  
137 a format suitable for spoken discourse (Rowley-Jolivet and Carter-Thomas, 2005)). Speech  
138 recognition can also be used to give feedback on pronunciation at the segmental level  
139 (Eskenazi, 1999; Neri et al., 2002; Hincks, 2003a). However, speech recognition for the pur-  
140 poses described here would need to be adapted to the individual speaker's voice, a time-con-  
141 suming process that does not lend itself to classroom applications (Coniam, 1999).

142 Fig. 1 illustrates how an automatic feedback mechanism could consist of two parallel  
143 processing operations, one conducted by the recognizer, and the other by speech analysis.  
144 This paper discusses the features in bold text, that is, pitch variation and speech rate. An  
145 appropriate and friendly feedback interface would be an animated face that could respond  
146 alertly to lively speech but would lose attention, perhaps even fall asleep, if the prosody  
147 failed to show any characteristics of liveliness. In the more distant future, a feedback  
148 mechanism could also incorporate a camera and software for processing speaker gaze, fa-  
149 cial expression and body language (Fig. 1).

150 An application of this kind places the computer in a supportive rather than a tutorial  
151 role (e.g., Levy, 1997). The system would not presume to correct the user, but merely act as  
152 a tool for quantifying the amount of prosodic variation. This allows the computer to do  
153 what computers have been proven to do well, which is to facilitate and support human  
154 communication, and avoids the pitfalls associated with the artificial intelligence required  
155 for tutorial systems.

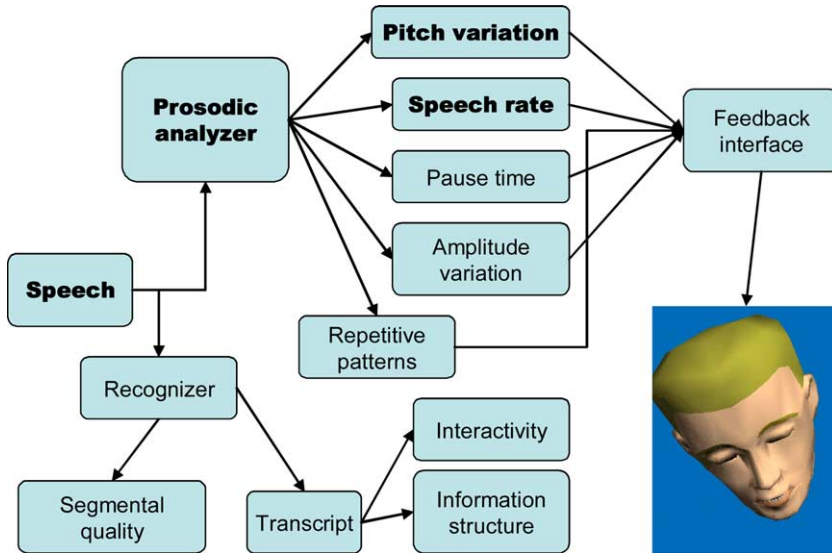


Fig. 1. Schematic design of an automatic feedback mechanism for public speaking. Features in bold are treated in this paper.

156 This investigation thus asks whether a metric created from data automatically derived  
 157 from speech analysis software can be used to describe the degree of speaker liveliness. The  
 158 hypothesis tested is that the higher the standard deviation in fundamental frequency, the  
 159 more a speech sample will be perceived as lively. This has been shown in earlier work using  
 160 synthesis of a single utterance (Traunmüller and Eriksson, 1995), but has not been tested  
 161 on naturally occurring speech.

## 162 2. Method

### 163 2.1. Material

164 The material under investigation comes from a corpus consisting of 35 ten-minute oral  
 165 presentations made by undergraduate engineering students, from four different courses  
 166 and three different proficiency levels of Technical English.<sup>3</sup> Many of these students expect  
 167 to find employment in companies whose official language is English, and find instruction  
 168 in English presentation skills to be a valuable part of their education. All students had ta-  
 169 ken an in-house placement test to determine their English proficiency, and all had studied  
 170 English for at least six years in the regular Swedish school system, where great emphasis is  
 171 placed on oral competence (Oscarson, 1995; Erickson, 2004). The students were audio re-  
 172 corded, with their written permission, in the classroom as they fulfilled a major, graded

<sup>3</sup> An analysis of the lexis and of the pronunciation errors found in part of this corpus was published as Hincks (2003b). That study found a very low frequency (less than 0.05% of words) of pronunciation errors that were likely to either impede intelligibility or be negatively perceived.

173 requirement in their courses in Technical English. The equipment used was a MiniDisc re-  
174 corder and a small clip-on microphone.

175 The prosodic analysis has been performed on a sub-corpus of 18 recordings. The crite-  
176 ria for inclusion in this sub-corpus were the student's sex, first language, and score on the  
177 in-house placement test in English. The goal was to have gender-balanced groups of six  
178 intermediate, six upper-intermediate, and six advanced students, all native speakers of  
179 Swedish. Because there were only eleven recordings of females (reflecting the gender bal-  
180 ance at a college of engineering), the males were chosen to match in score the nine females  
181 who met the score requirements. The mean age of the students was 25.5, SD 2.6. The in-  
182 house placement test had a maximum score of 100 and measured ability in writing, gram-  
183 mar and vocabulary. Group A had scores from 50 to 60, Group B from 61 to 70, and  
184 Group C from 80 to 90. For the purposes of comparison with other student groups, it  
185 is useful to provide estimations of student competence in the Council of Europe's Com-  
186 mon Framework of Reference (Council, 2001). The students in groups A and B could  
187 probably be placed in Council group B2, and the students in group C in the upper ranges  
188 of Council group C1.

189 The presentations were transcribed using regular English orthography, and the sound  
190 files digitally stored using 16 kHz sampling. In preparation for prosodic analysis, interrup-  
191 tions in the presentations due to equipment problems, pauses of 10 s or more, or question-  
192 and-answer segments were edited away.

## 193 2.2. Pitch extraction and variation

194 To enable smoother handling of the large quantity of data, the 7–10-min long record-  
195 ings were divided into 30-s segments stored as separate sound files. Each file was then pro-  
196 cessed using the speech analysis function of the program WaveSurfer (Sjölander and  
197 Beskow, 2000). WaveSurfer was configured to search for pitch (fundamental frequency)  
198 at between 60 and 400 Hz for the male voices and between 120 and 500 Hz for the female  
199 voices. The pitch contour produced for each file was visually inspected for evidence of  
200 miss-readings, and the location of these errors noted. WaveSurfer extracts a pitch value  
201 for every 10 ms of speech; that is, 100 values per second. These values were imported into  
202 a spreadsheet program for further analysis.

203 The next steps in the analysis were to delete from the spreadsheet program all values of  
204 zero (from unvoiced segments or silence), and all values that corresponded to errors in the  
205 pitch extraction (as evidenced by the visual inspection). Then, for each 10 s of speech, the  
206 mean and standard deviations of the pitch were calculated. Ten seconds of speech was cho-  
207 sen as a good unit for data analysis because it was enough time to guarantee the inclusion  
208 of a fair amount of speech at normal pausing rates.

209 The raw standard deviation of the pitch is unsuitable in itself as a measure of pitch vari-  
210 ation (Traunmüller and Eriksson, 1995). This is because of the differences among speak-  
211 ers, and particularly between sexes, of the mean pitch level. The higher the frequency of  
212 our voices, the larger the standard deviation will be in normal speech, and this would give  
213 an 'unfair' advantage to female speakers over male. Therefore, in order to make valid com-  
214 parisons among speakers, the standard deviation is expressed as a percentage of the mean.  
215 For example, a standard deviation of 21 and a mean frequency of 115 (a male voice) yields  
216 a value of 0.183; a standard deviation of 36 and a mean frequency of 195 (a female voice)

217 yields a value of 0.185. To simplify expression, I have termed this value the PVQ, for pitch  
218 variation quotient.

219 With this method, the PVQ was calculated for every 10 s of speech for up to 10 min of  
220 speech for each of the 18 students. This yielded a database of 986 values for the entire sub-  
221 corpus.

### 222 2.3. Perception test

223 A perception test was prepared to test the hypothesis that speech with higher PVQ  
224 would be perceived as livelier speech. Nine 10-s samples of speech per speaker, represent-  
225 ing the speaker's three lowest PVQs, three mean PVQs, and three peak PVQs, were se-  
226 lected as test files. If any of these nine files contained a pause that was longer than 4 s,  
227 it was substituted by another for two reasons: one, that the PVQ value was less stable  
228 when it represented less speech, and two, that judges should not be rating a speech sample  
229 that consisted of nearly 50% silence. Each speaker was thus represented by one and a half  
230 minutes of speech in nine separate test files, giving 81 test files to rate for each sex.

231 Separate tests were prepared for male and female speakers using Visor from the Spruce  
232 package (Granqvist, 2003). Respondents were presented with a randomized collection of  
233 test file icons on a computer screen and were instructed to listen to each file and then move  
234 it to an undivided scale whose endpoints were 'lively' and 'monotone'. They could listen to  
235 the files as many times as they wanted to and move them as many times as they wished.  
236 They were instructed to disregard impressions of speaker proficiency and focus on quali-  
237 ties of engagement and liveliness, but no judge was aware that it was pitch variation that  
238 was being tested. The tests took between one half hour and one and one half hours to com-  
239 plete, and respondents completed the male and female tests on different days, in random-  
240 ized order. The respondents were eight (two male) university teachers of English, who were  
241 natives of Britain (2), Sweden (2), USA, Brazil, Germany and Turkey. None of the teach-  
242 ers were specialists in teaching pronunciation, but three of them had extensive experience  
243 in teaching presentation skills.

### 244 2.4. Speaking rate

245 Speaking rate is often expressed in words per minute (WPM). This is an imprecise  
246 measurement for a number of reasons (Griffiths, 1990, 1991; Griffiths and Beretta,  
247 1991; Kormos and Dénes, 2004). In contrast, expressing speaking rate in syllables per sec-  
248 ond (SPS) gives a number of advantages. First, it provides a fair comparison between  
249 speakers who use long words versus those who use shorter words. Second, it allows  
250 cross-linguistic comparisons between languages with different average word length. Third,  
251 it provides a more local measurement so that variations in speaking rate can be tracked.  
252 Finally, to calculate WPM one needs a transcript of the event. Since syllables can be char-  
253 acterized as bursts of acoustic energy corresponding to the syllabic nucleus, their number  
254 can be counted, or at least reliably estimated, on the basis of the speech waveform without  
255 knowing what has been said. For this research, however, a manual rather than automatic  
256 method was used for calculating SPS.

257 Another relevant variable, particularly when analyzing L2 speech, is mean length of  
258 runs (MLR). This is the number of syllables the speaker has uttered between pauses. In

259 this paper, as in Kormos and Dénes (2004), a pause is defined as a silent interval longer  
260 than 250 ms, or one quarter of a second.

261 Speaking rates in WPM and SPS, as well as MLR were calculated for the entire presen-  
262 tation of each of the 18 speakers.

### 263 3. Results

264 There are thus a number of variables to take into consideration. For each speaker, there  
265 is a value representing proficiency in English, a value for mean pitch variation throughout  
266 the presentation, values for mean speech rate and mean length of runs, and finally the  
267 mean of the nine liveliness ratings per speaker. These values can be used to characterize  
268 the speakers and their presentations. In addition, the liveliness ratings for 162 individual  
269 10-s samples of speech can be seen in relation to PVQ, speech rate, and mean length of run  
270 within that sample.

#### 271 3.1. Pitch variation results

272 PVQ values for all individual 10-s samples in the corpus are shown in Fig. 2. For both  
273 males and females, the lowest values in the corpus are about 0.06 and the values follow  
274 each other nearly exactly up to 0.157, where they diverge, with the male values higher than  
275 the female. The maximum values for males are above 0.30, while female values reach just  
276 above 0.25. (Fig. 2)

277 Fig. 3 is an example of the PVQ development over the whole presentation of two male  
278 speakers from group C. The speech of speaker 14, whose values vary greatly with 0.23 as  
279 the mean, was described by his teacher in written comment on the presentation as being

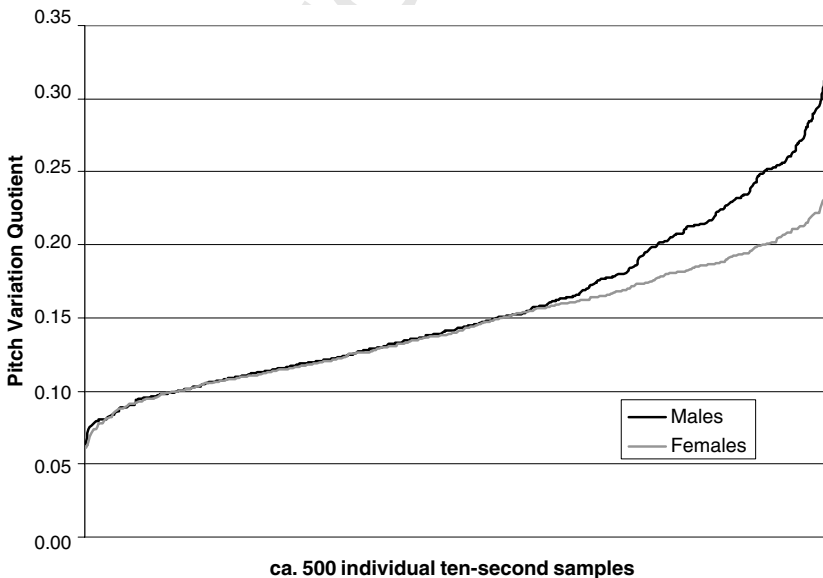


Fig. 2. Distribution of all PVQ values for all 10-s samples in presentation corpus. 492 samples for males and 494 for females.

280 “well-modulated” and with “varied intonation.” In contrast, the speech of speaker 17,  
 281 whose values range mostly between 0.10 and 0.15, was described as being “a little  
 282 deadpan.”(Fig. 3)

283 The relationship between the speakers’ proficiency in English and mean pitch variation  
 284 in the presentation as a whole is shown in Fig. 4, which plots mean PVQ per speaker  
 285 against the speaker’s score on the in-house placement test. For female speakers, there is  
 286 a significant correlation between pitch variation and proficiency ( $r = .83, n = 9, p < .01$ ),  
 287 but for the males the relationship is not significant. (Fig. 4)

### 288 3.2. Perception test results

#### 289 3.2.1. Inter-rater reliability

290 A reliability analysis performed on the results of the perception test gave high values for  
 291 Cronbach’s alpha: .98 and .95 for the composite judgments of male and female speakers,  
 292 respectively, and .94 and .90 for male and female speech samples.

293 Many of the judges commented that they found the liveliness rating task easier for the  
 294 male speech than for the female speech, and this is shown in the results of the perception  
 295 test. Table 1 shows the correlations between each judge’s liveliness ratings per speaker (the  
 296 mean of the ratings of the nine samples per speaker) and the means of the PVQs of the  
 297 speech that was rated. The table also shows the correlations between the liveliness ratings  
 298 per speaker and the speaker’s score on the English proficiency test. The ratings of male  
 299 speakers reach a higher level of correlation with PVQ than the ratings of female speakers  
 300 do. The differences between perceptions of males and females are even more striking when  
 301 it comes to the correlations with the student’s score on the English proficiency test. The  
 302 judges appear to have succeeded with the instruction to ignore questions of proficiency  
 303 when rating the male speakers, since for all judges correlations with proficiency are lower

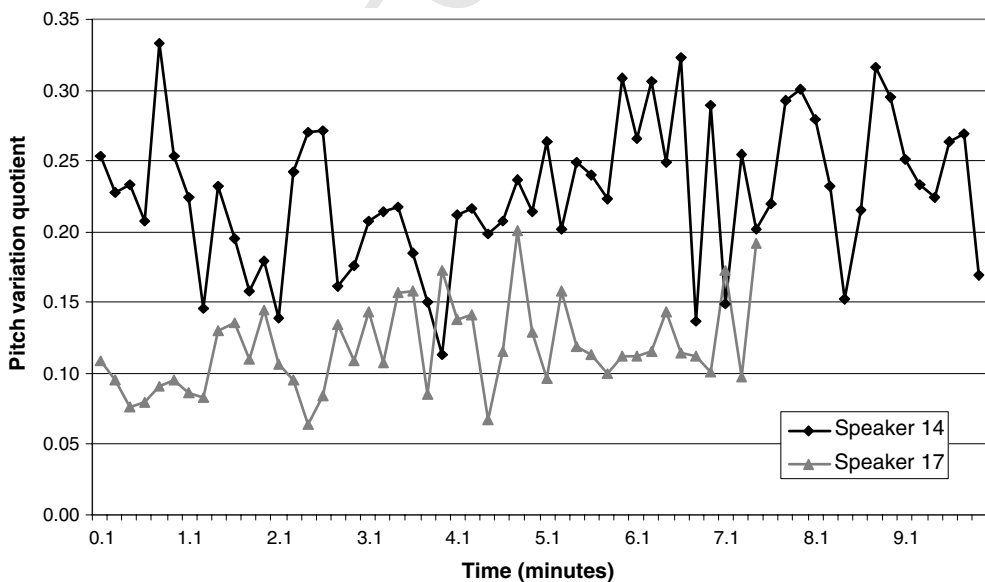


Fig. 3. The development of PVQ over time for two males from Group C.

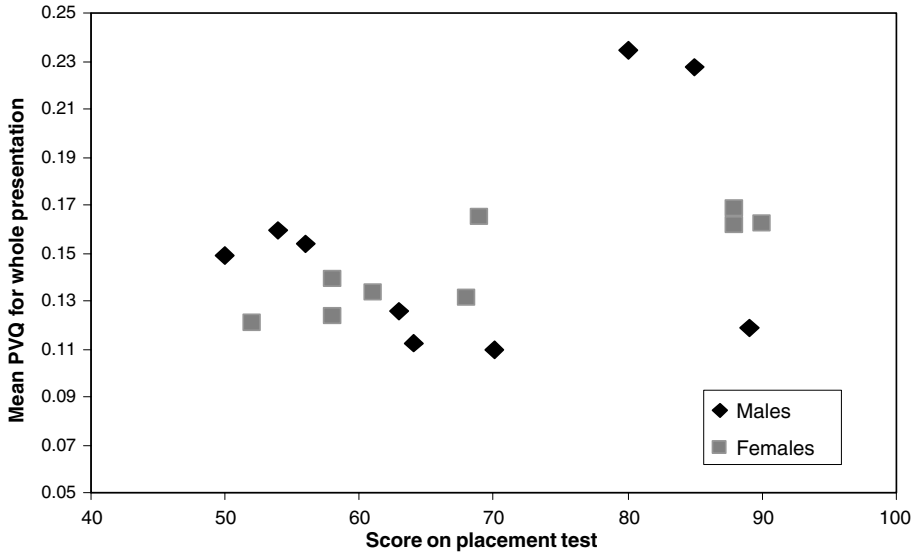


Fig. 4. Mean pitch variation quotient per speaker in relation to proficiency in English as measured by in-house placement test.

Table 1

Pearson correlations between liveliness ratings of speakers and both PVQ and English proficiency as measured by placement test, by judge

Judge	Pitch variation		English proficiency	
	Males	Females	Males	Females
1	.86**	.77**	.47	.87**
2	.95**	.75**	.40	.84**
3	.88**	.38	.52	.57
4	.91**	.68*	.41	.72*
5	.87**	.61*	.63*	.71*
6	.88**	.69*	.59*	.80*
7	.68*	.59*	.32	.56
8	.75*	.67*	.60*	.83**

n = 9 males, 9 females.

\* p < .05, one tail.

\*\* p < .01, one tail.

304 than with pitch variation. The reverse is true for the ratings of female speakers, where for  
 305 seven of the judges, correlations with proficiency are higher than with pitch variation. This  
 306 could be partly due to the fact that for the females in this corpus there was also a stronger  
 307 relationship between PVQ and English proficiency, as shown in Fig. 4 (Table 1).

308 3.2.2. Perceptions of speakers

309 Fig. 5 plots means of the PVQs per speaker against the means of the liveliness ratings of  
 310 all judges per speaker. The program used to produce the perception test uses a visual scale  
 311 that transforms the placement of an icon on a screen to values between 0 and 1000, which  
 312 is the scale used here on the x-axis. Males are shown with filled symbols and females with

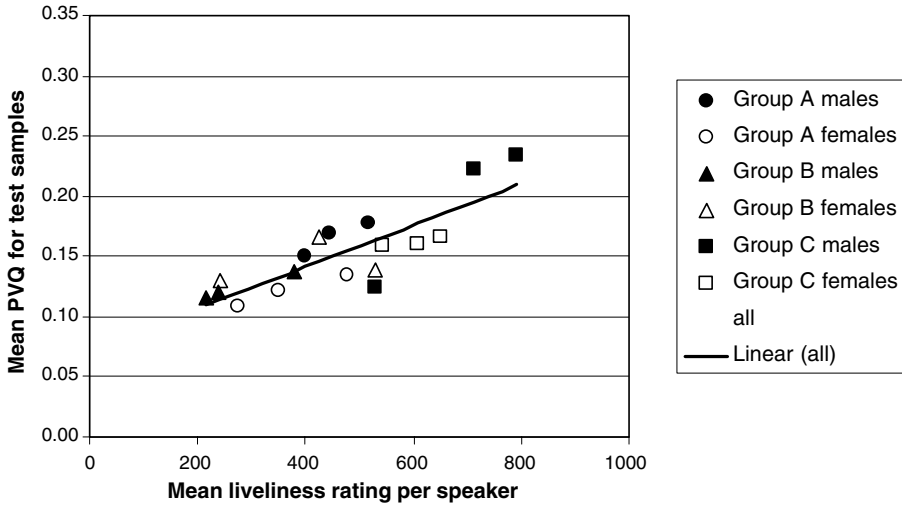


Fig. 5. Mean pitch variation quotient for all test samples per speaker plotted against mean liveliness rating per speaker.

313 unfilled symbols, with different shapes for the different groups. The correlation between  
 314 the values is significant ( $n = 18, p < .01$ ) and strong:  $r = .83$ , indicating that the PVQ met-  
 315 ric is a reliable indicator of speaker liveliness (Fig. 5).

### 316 3.2.3. Perceptions of speech samples

317 The composite of the liveliness ratings for the nine speech samples per speaker gives a  
 318 more reliable characterization of speakers than what can be discerned from the perception  
 319 of liveliness in a single 10-s sample of speech, but even at this level the correlations are also  
 320 significant ( $n = 81, p < .01$ ) for both sexes. The mean liveliness ratings for all individual  
 321 test files are plotted against PVQ in Fig. 6.<sup>4</sup> Once again correlations for males are higher  
 322 than for females: .70 for the males and .64 for the females. Generally, speech samples with  
 323 low PVQ occupy the lower left quadrant and samples with high PVQ occupy the upper  
 324 right quadrant, most noticeably a group of six samples from males of Group C. The lower  
 325 right quadrant contains mostly squares (the most proficient group), indicating that judges  
 326 perceived them as lively even when their PVQs were not high. In contrast, speakers from  
 327 groups A and B occupy the upper left quadrant, indicating that high pitch variation was  
 328 not always perceived as liveliness for these less fluent speakers (Fig. 6).

### 329 3.3. Temporal measurements

330 Table 2 shows means and standard deviations per student proficiency level for three dif-  
 331 ferent temporal measures: mean length of runs (MLR), words per minute (WPM), and syl-  
 332 lables per second (SPS). The advanced students, Group C, spoke more quickly and  
 333 produced more speech between pauses than the two intermediate groups. The mean tem-

<sup>4</sup> One of the female test files with an outlying position was re-examined, was found to have violated the pause length criterion, and its results were removed. Another two female test files were discovered to be duplicates and the results for one of them removed.

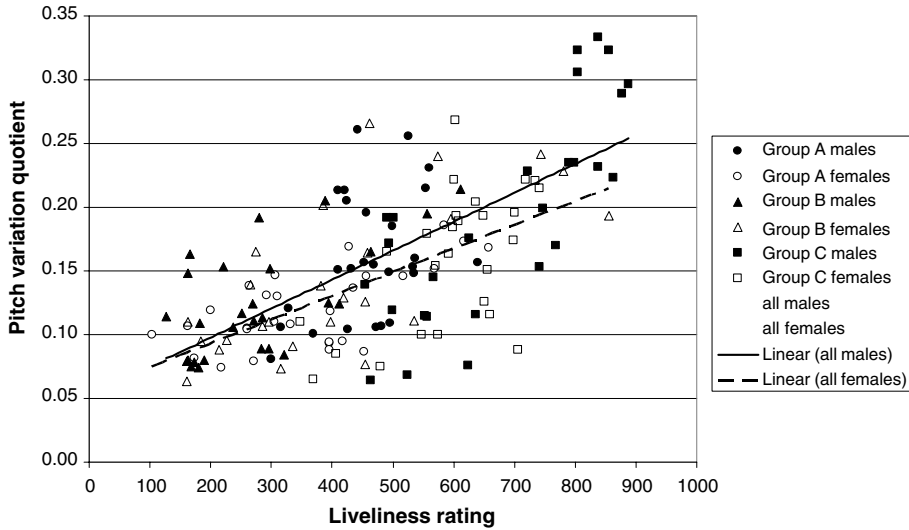


Fig. 6. Pitch variation quotient for 160 10-s samples of speech versus mean liveliness rating from panel of eight judges.

334 poral measures correlate perfectly with each group’s mean score on the written placement  
 335 test, supporting earlier findings on the relationship between speaker proficiency and speak-  
 336 ing rate (Towell et al., 1996; Cucchiariini et al., 2000; Kormos and Dénes, 2004) (Table 2).  
 337 The mean results for MLR, WPM and SPS presented in Table 2 have been calculated  
 338 on nearly an hour of speech per student group. In addition, local values were calculated  
 339 for each of the 160 ten-second test samples used in the perception test, in order to examine  
 340 the extent to which the liveliness ratings correlated with rate of speech. Table 3 shows cor-  
 341 relations between the liveliness rating and two temporal measures as well as with PVQ.  
 342 These results reveal interesting differences between the raters’ perceptions of liveliness in  
 343 male and female speakers. While both temporal measures had no correlation with the per-  
 344 ception of liveliness in male speakers, the correlation between MLR and liveliness in the  
 345 female test samples was .72, higher than the correlation of liveliness with PVQ, which  
 346 for females was only .64 for individual samples. Plain speaking rate as measured in SPS  
 347 had no effect on liveliness judgments for either sex (Table 3).

348 **4. Discussion**

349 The PVQ variable is a very good though imperfect correlate of the perception of live-  
 350 liness in student presentation speech. The following sections look at the factors that could  
 351 contribute to moderating the correlations between perceived liveliness and PVQ.

Table 2  
 Mean values of temporal measures for each student group

Group	n	Placement test mean (SD)	MLR mean (SD)	WPM mean (SD)	SPS mean (SD)
A	6	54.7 (3.3)	7.10 (1.34)	115 (11)	2.75 (.25)
B	6	65.8 (3.7)	8.02 (1.11)	118 (11)	2.86 (.26)
C	6	86.7 (3.7)	9.80 (2.12)	128 (18)	3.14 (.36)

Table 3

Pearson correlation coefficients between three prosodic variables and liveliness rating for individual speech samples

	SPS	MLR	PVQ
Males	.06	-.06	.70**
Females	.16	.72**	.64**

$n = 81$  male samples, 79 female samples.

\*\*  $p < .01$ .

#### 352 4.1. Causes of high pitch variation

353 Many of the high PVQ files were examples of speakers successfully using pitch contrasts  
 354 to structure the content of their presentation when introducing items in a series or moving  
 355 from an old topic to a new one. High PVQ was also evident when speakers chose to illus-  
 356 trate points on a blackboard, and of course when they showed extra enthusiasm for their  
 357 topic. Samples of this nature generally received correspondingly high liveliness ratings. On  
 358 the other hand, some of the samples with high PVQ were due to speakers using large pitch  
 359 variation for other reasons. For example, pitch resets caused by disfluencies or nervous-  
 360 ness could lead to high PVQ values. A few of the speakers, most of them female, had a  
 361 habit of speaking with high rises at the end of an utterance,<sup>5</sup> which would contribute to  
 362 a high PVQ though the speech could well be interpreted as uncertain. All of the speakers  
 363 bore traces of Swedish prosodic patterns to one degree or another, though this was most  
 364 evident in the lower-proficiency speakers. Because Swedish has distinctive word accents  
 365 signaled by pitch movement within one word, the presence of these patterns could contrib-  
 366 ute to a high variation in pitch that was not necessarily rated as lively by the judges. Fur-  
 367 thermore, some of the samples with high PVQs contained Swedish names uttered as they  
 368 should be intoned in Swedish (with large pitch movement); neither would these files receive  
 369 high liveliness ratings. It is likely that some of the high PVQ files were perceived as being  
 370 accented rather than lively, though experimental tests would be required to confirm this  
 371 observation.

#### 372 4.2. Causes of high liveliness ratings

373 The five test files that received mean liveliness ratings of above 600 and yet had PVQ  
 374 below 0.13, found in the lower right quadrant of Fig. 6, all came from the highly proficient  
 375 speakers of Group C. Four of these files had been selected for the test as examples of that  
 376 speaker's lowest PVQs; in giving them relatively high ratings, the judges may have been  
 377 responding to a sort of 'halo effect' where they rated a file highly because they had rated  
 378 other files from that speaker highly.

#### 379 4.3. Sex differences

380 Below the PVQ value 0.157, males and females in the corpus show a nearly identical  
 381 distribution of values (Fig. 2). Above this point, the values diverge, with males showing

<sup>5</sup> Unlike North American 'uptalk,' however, these high rises found in some Swedish speakers do not convey the impression that a declarative is a question. The rise is more like a plateau.

382 more pitch variation than females. Interestingly, this point appears again in Fig. 6, where  
383 all the PVQ values are plotted against the mean of the liveliness ratings given to them by  
384 the panel of eight judges. The intersection of the mid-value 500 with the midpoint between  
385 the regression lines of the males and females is very close to the same point at which the  
386 male and female values diverge. The judges are indicating that PVQ values higher than  
387 0.157 are perceived as being on the upper end of the lively-monotone scale, and the data-  
388 base indicates that males had a better ability than females to achieve these higher values in  
389 their speech. Furthermore, Fig. 6 shows that the highest male PVQ values received corre-  
390 spondingly high liveliness ratings, while the highest female PVQ values did not receive the  
391 highest liveliness ratings. The data shown in Table 3 also indicate that males and females  
392 may differ in the production and perception of liveliness. For 81 test files, male liveliness  
393 ratings show no correlation whatsoever with the fluency of the speech, as measured in  
394 mean length of runs, or how many syllables the speakers uttered between pauses  
395 >250 ms. Female liveliness ratings, on the other hand, correlate moderately with MLR,  
396 more strongly than with the PVQ variable. In other words, the raters may have judged  
397 the female speakers on how fluent they were, but not the males.

398 These possible differences in the perception and production of liveliness in male and fe-  
399 male speech are an unexpected result. Traunmüller and Eriksson (1995) found no effect of  
400 speaker identity on liveliness perceptions in their study of synthetic speech. However, stud-  
401 ies of natural speech have drawn different conclusions. Aronovich (1976) and Henton  
402 (1989) concluded that we expect male speech to lack variability and female speech to have  
403 a lot of variability; therefore, when males do use a lot of variability, its effect is very salient.  
404 The database in this study consists of six speakers per proficiency group but only three of  
405 each sex. If different conclusions need to be drawn for the males and females, the groups in  
406 this investigation have become too small to say anything conclusive. It is possible that,  
407 finding the rating task difficult for females, the judges sorted the females according to per-  
408 ceived proficiency in English.

#### 409 4.4. *Other variables*

410 This paper has looked at variables of pitch and tempo, but speakers have a third means  
411 available for varying intonation, and that is intensity, or loudness. It may be that some of  
412 the speakers who were rated as lively though their pitch variation was low were using  
413 amplitude variation effectively. However, in order to gather reliable measurements of  
414 loudness, speakers must be recorded under much stricter conditions, for example, using  
415 a head-mounted microphone whose distance from the mouth is kept constant.

416 The results presented above would be simpler to interpret if the database was of native  
417 speech where issues of fluency and foreign accent are moot. Researchers with access to  
418 such a database are encouraged to test these methods on that speech. It is likely that  
419 the correlations between pitch variation and perceived liveliness would be even greater  
420 in such investigations.

### 421 5. Conclusion and pedagogical implications

422 To conclude, I would like to draw some preliminary conclusions regarding the pitch  
423 variation levels that correspond to lively versus monotone speech. The automatic feedback  
424 mechanism described in Fig. 1 could be configured to respond negatively to PVQ values

425 that were under 0.15 for lengthy periods of time. Speakers could be encouraged to hold  
426 mean values between 0.15 and 0.25, and be rewarded for the occasional peak above  
427 0.25. The level of liveliness could be adapted to the speaking genre; while one level would  
428 be suitable for an evangelist, another is clearly more appropriate to an academic confer-  
429 ence presentation. Furthermore, people's perceptions of what is appropriate and pleasing  
430 may be individually and culturally determined to one extent or another.

431 In terms of speaking rate, a reasonable approach could be for speakers of a given pro-  
432 ficiency to aspire to the speaking rate of the next proficiency level. The intermediate  
433 (groups A and B) speakers could aspire to reach speaking rates above three syllables  
434 per second. Clearly there is a maximum speaking rate above which comprehension be-  
435 comes impaired, particularly when English is being used in international settings ([Camic-](#)  
436 [iottoli, 2005](#)). Some native speakers may be unaware of what kinds of speaking rates may  
437 be inappropriate when a majority of the audience consists of non-natives who, though flu-  
438 ent speakers, may not be able to process content at the same speed as natives ([Griffiths and](#)  
439 [Beretta, 1991](#)). These speakers may benefit from feedback telling them to slow down their  
440 delivery. Furthermore, since problems with monotonous speech delivery are not restricted  
441 to non-native speakers, it is likely that both native and non-native speakers would find the  
442 pitch variation feedback useful.

443 [Hahn \(2004\)](#) has shown that speaking with sentence focus is important for the success-  
444 ful communication of content, and [Pickering \(2004\)](#) has shown that non-natives have a  
445 harder time than natives in modifying their pitch at the discourse level. Pickering's non-  
446 native subjects were native speakers of Mandarin; other research ([Wennerstrom, 1994](#))  
447 has shown that native speakers of a European language do not exhibit the same difficulties  
448 as speakers of Asian languages in using pitch to structure discourse in English. Swedish is  
449 a language with a close genetic relationship to English, and can be characterized, like Eng-  
450 lish, as a stress-timed rather than a syllable-timed language. As in English, new informa-  
451 tion in an utterance should receive more focus than given information. This focus is  
452 achieved primarily by pitch movement through the focused word, usually accompanied  
453 by lengthening of the focused syllable and an increase in intensity. The Swedish speakers  
454 in this study should thus not be experiencing negative transfer from their L1 when it comes  
455 to providing focus in the appropriate parts of an utterance or in making pitch resets when  
456 introducing a new topic, and there was no evidence in the corpus that they had these prob-  
457 lems. Yet still some speakers were more monotone than others. It is likely that affect – ner-  
458 vousness – is a large reason for this. It stands to reason that speakers who are nervous  
459 presenting in their own language are even more nervous presenting in a second language.  
460 These speakers need simply to be reminded not to forget about intonation as they speak –  
461 hence the admonitions in the public speaking manuals. A major benefit of an automatic  
462 feedback mechanism would be simply to help speakers notice problems and to track their  
463 own improvement. When a speaker has learned to let pitch movement loose on the impor-  
464 tant parts of his or her message, the result should be livelier speech and better  
465 communication.

## 466 Acknowledgements

467 Many thanks to the colleagues who generously contributed their time evaluating the  
468 test files, the students who allowed me to record them, and David House and Björn Gran-  
469 ström. This work was jointly funded by the Unit for Language and Communication and

470 the Centre for Speech Technology at KTH. Portions of this paper were presented at the  
471 InSTILL/ICALL Conference in Venice, June 2004.

## 472 References

- 473 Anderson-Hsieh, J., 1992. Interpreting visual feedback on suprasegmentals in computer assisted pronunciation  
474 instruction. *CALICO Journal* 11 (4), 5–21.
- 475 Aronovich, C.D., 1976. The voice of personality: stereotyped judgements and their relation to voice quality and  
476 sex of speaker. *Journal of Social Psychology* 99, 207–220.
- 477 Bot, K.d., Mailfert, K., 1982. The teaching of intonation: fundamental research and classroom applications.  
478 *TESOL Quarterly* 16 (1), 71–77.
- 479 Brazil, D., 1997. *The Communicative Value of Intonation in English*. Cambridge University Press, Cambridge.
- 480 Camiciottoli, B., 2003. Interactive discourse structuring in L2 guest lectures: some insights from a comparative  
481 corpus-based study. *Journal of English for Academic Purposes* 3 (1), 39–54.
- 482 Camiciottoli, B., 2005. Adjusting a business lecture for an international audience: a case study. *English for*  
483 *Specific Purposes* 24, 183–199.
- 484 Coniam, D., 1999. Voice recognition software accuracy with second language speakers of English. *System* 27 (1),  
485 49–64.
- 486 Council, E., 2001. *Common European Framework of Reference for Languages*. Cambridge University Press,  
487 Cambridge.
- 488 Cucchiaroni, C., Strik, H., Boves, L., 2000. Different aspects of expert pronunciation quality ratings and their  
489 relation to scores produced by speech recognition algorithms. *Speech Communication* 30, 109–119.
- 490 Erickson, G., 2004. English: here and there and everywhere. En undersökning av ungdomars kunskaper i och  
491 uppfattningar om engelska i åtta europeiska länder. Stockholm, Skolverket: 90..
- 492 Eskenazi, M., 1999. Using automatic speech processing for foreign language pronunciation tutoring: Some issues  
493 and a prototype. *Language Learning and Technology* 2 (2), 62–76.
- 494 Grandstaff, D., 2004. *Speaking as a Professional*. W.W. Norton & Co.
- 495 Granqvist, S., 2003. *Computer methods for voice analysis*. Department of Speech, Music and Hearing.  
496 Stockholm, KTH. PhD. thesis..
- 497 Griffiths, R., 1990. Speech rate and NNS comprehension: a preliminary study in time-benefit analysis. *Language*  
498 *Learning* 40 (3), 311–336.
- 499 Griffiths, R., 1991. Pausological research in an L2 context: a rationale and review of selected studies. *Applied*  
500 *Linguistics* 12 (4), 345–364.
- 501 Griffiths, R., Beretta, A., 1991. A controlled study of temporal variables in NS-NNS lectures. *RELJ Journal* 22  
502 (1), 1–19.
- 503 Hahn, L.D., 2004. Primary stress and intelligibility: research to motivate the teaching of suprasegmentals. *TESOL*  
504 *Quarterly* 38 (2), 201–223.
- 505 Hardison, D., 2004. Generalization of computer-assisted prosody training: quantitative and qualitative findings.  
506 *Language Learning and Technology* 8 (1), 34–52.
- 507 Henton, C., 1989. Fact and fiction in the description of female and male pitch. *Language and Communication* 9  
508 (4), 299–311.
- 509 Hincks, R., 2003a. Speech technologies for pronunciation feedback and evaluation. *ReCALL* 15 (1), 3–20.
- 510 Hincks, R., 2003b. Pronouncing the Academic Word List: Features of L2 student oral presentations. In: 15th  
511 International Congress of Phonetic Sciences, Barcelona, ICPhS Organizing Committee.
- 512 Jenkins, J., 2000. *The Phonology of English as an International Language: New Models, New Norms, New*  
513 *Goals*. Oxford University Press, Oxford.
- 514 Kormos, J., Dénes, M., 2004. Exploring measures and perceptions of fluency in the speech of second language  
515 learners. *System* 32, 145–164.
- 516 Lamerton, J., 2001. *Collins Complete Guide to Public Speaking*. HarperCollins.
- 517 Levis, J., Pickering, L., 2004. Teaching intonation in discourse using speech visualization technology. *System* 32,  
518 505–524.
- 519 Levy, M., 1997. *Computer-Assisted Language Learning*. Clarendon Press, Oxford.
- 520 Molholt, G., 1988. Computer-assisted instruction in pronunciation for Chinese speakers of American English.  
521 *TESOL Quarterly* 22 (1), 91–111.

- 522 Morell, T., 2004. Interactive lecture discourse for university EFL students. *English for Specific Purposes* 23, 325–  
523 338.
- 524 Neri, A., Cucchiari, C., Strik, H., Boves, L., 2002. The pedagogy-technology interface in computer assisted  
525 pronunciation training. *Computer-Assisted Language Learning* 15 (5).
- 526 Oscarson, M., 1995. A national evaluation programme in the Swedish compulsory school: assessment of  
527 achievement in foreign languages. *System* 23 (3), 295–306.
- 528 Öster, A.-M., 1998. Spoken L2 teaching with contrastive visual and auditory feedback. In: *Proceedings of ICSLP*,  
529 Sydney.
- 530 Pickering, L., 2004. The structure and function of intonational paragraphs in native and nonnative speaker  
531 instructional discourse. *English for Specific Purposes* 23, 19–43.
- 532 Rowley-Jolivet, E., Carter-Thomas, S., 2005. Genre awareness and rhetorical appropriacy: manipulation of  
533 information structure in the international conference setting. *English for Specific Purposes* 24, 41–64.
- 534 Simpson, R.C., Lee, D.Y.W., Leicher, S., 2003. MICASE Manual: The Michigan Corpus of Spoken Academic  
535 English. English Language Institute, The University of Michigan, Ann Arbor, MI, USA.
- 536 Sjölander, K., Beskow, J., 2000. WaveSurfer: An open source speech tool. ICSLP 2000, Available from: [http://  
537 www.speech.kth.se/snack/](http://www.speech.kth.se/snack/).
- 538 Towell, R., Hawkins, R., Bazergui, N., 1996. The development of fluency in advanced learners of French. *Applied*  
539 *Linguistics* 17 (1), 84–119.
- 540 Traunmüller, H., Eriksson, A., 1995. The perceptual evaluation of  $F_0$  excursions in speech as evidenced in  
541 liveliness estimations. *Journal of the Acoustical Society of America* 97 (3), 1905–1915.
- 542 Wennerstrom, A., 1994. Intonational meaning in English discourse. *Applied Linguistics* 15, 399–421.
- 543 Wennerstrom, A., 2001. *The Music of Everyday Speech*. Oxford University Press, New York.
- 544 Ventola, E., Shalom, C., Thompson, S. (Eds.), 2002. *The Language of Conferencing*. Peter Lang, Frankfurt am  
545 Maim.
- 546