

Title	Acoustic features correlated to perceived urgency in evacuation announcements
Author(s)	Kobayashi, Maori; Hamada, Yasuhiro; Akagi, Masato
Citation	Speech Communication, 139: 22-34
Issue Date	2022-03-06
Type	Journal Article
Text version	author
URL	http://hdl.handle.net/10119/18823
Rights	Copyright (C)2022, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (CC BY-NC-ND 4.0). [http://creativecommons.org/licenses/by-nc-nd/4.0/] NOTICE: This is the author's version of a work accepted for publication by Elsevier. Maori Kobayashi, Yasuhiro Hamada, Masato Akagi, Speech Communication 139, 2022, 22-34.
Description	

1 Title:
2 Acoustic features correlated to perceived urgency in evacuation
3 announcements
4
5 Authors:
6 Maori Kobayashi, Yasuhiro Hamada, and Masato Akagi
7
8 Affiliation:
9 School of Information Science
10 Japan Advanced Institute of Science and Technology
11 1-1 Asahi-dai, Nomi, Ishikawa, Japan
12
13 Corresponding Author:
14 Maori Kobayashi, Ph.D.
15 E-mail: maori-k@jaist.ac.jp

16 Abstract:

17 To encourage prompt evacuation behavior during disasters, evacuation
18 announcement systems are required to adjust the perceived urgency of
19 announcements to the danger of a situation while maintaining voice clarity. In
20 this study, we aimed to understand acoustic features correlated to the perceived
21 urgency in speech when clarity is maintained. For this purpose, we used a speech
22 synthesis tool to manipulate the key acoustic features: the time duration, F0
23 (instantaneous fundamental frequency), and spectrum. Specifically, we replaced
24 these features in five real evacuation announcements that were spoken clearly by
25 a TV announcer during an actual disaster and had different magnitudes of
26 urgency. We found quantitatively and qualitatively that the perceived urgency
27 was mostly influenced by the F0 of speech. Furthermore, by manipulating the F0
28 time average and variation, we compared the influence of the F0's static constant
29 feature and its dynamic fluctuation pattern on the magnitude of perceived
30 urgency. The results indicated that both types of F0 information influenced the
31 magnitude of perceived urgency in real Japanese speech. Our results suggest that
32 the sense of urgency in evacuation announcements can be controlled by adjusting
33 the F0 while maintaining voice clarity.

34

35 Keywords:

36 vocoded speech, evacuation announcement speech, perceived urgency

37 1. Introduction

38 Japan faces the persistent threat of natural disasters and takes pains to
39 be prepared for the worst. An example of such preparation is a disaster
40 administration wireless communication system called *bousai-musen*, which
41 consists of a local warning system network. Many municipalities have banks of
42 loudspeakers mounted on poles as part of a broadcast system that stands ready
43 to alert residents to impending natural disasters or other large-scale civil
44 emergencies. The *bousai-musen* system uses speech and alarm sounds to raise
45 warnings of earthquakes or provide other vital information in an emergency,
46 giving residents valuable seconds to find a safe place.

47 Although the *bousai-musen* system has proven invaluable in
48 emergencies, problems have also been pointed out. For example, in the 2011
49 Great Tohoku Earthquake in Japan, it was reported that some citizens ignored
50 warnings and did not evacuate, despite hearing announcements calling for
51 evacuation. This phenomenon is considered to be due to normalcy bias, which
52 refers to a mental state that causes some people facing a disaster to underestimate
53 both the possibility of a disaster occurring and its possible effects (Drabek, 1986;
54 Omer and Alon, 1994).

55 In the case of the 2011 Great Tohoku Earthquake, it is assumed that
56 some people did not recognize the situation's danger because the *bousai-musen*
57 announcements were calm and polite. In contrast, fewer people fell victim to the
58 resulting tsunami in municipalities where the announcements sounded urgent
59 (Reports on tsunami evacuation measures, 2012). These reports indicate the need
60 to examine speech styles for conveying danger to people appropriately. It was
61 also reported that some people who read the evacuation announcements were
62 late to escape the tsunami (Reports on tsunami evacuation measures, 2012).
63 Accordingly, an automatic evacuation announcement system is necessary, and it
64 should control the magnitude of an announcement's urgency according to the
65 situation.

66 Previous studies have examined the speech duration, including the
67 speech rate (syllables per minute) and time intervals between sentences, the
68 speaker's gender, and the fundamental frequency (F0) as factors affecting the
69 perceived urgency in speech (Jang, 2007; Hellier et al., 2002; Park and Jang, 1999).

70 Notably, it was demonstrated that the time duration is essential for determining
71 the magnitude of perceived urgency: a reduction in duration increases the
72 magnitude (Hellier et al., 2002; Park and Jang, 1999; Jang, 2007). It was also
73 reported that women's speech gives the impression of higher urgency than men's
74 speech, which might result from differences in acoustic effects due to the
75 differing vocal-tract characteristics between genders (Edworthy et al., 1991;
76 Edworthy, 1994). Consistently with those findings, some reports showed that the
77 magnitude of perceived urgency is also influenced by the F0 (F0 average, F0
78 contour) of a sentence (Park and Jang, 1999). Also, in general, the louder a
79 warning signal, the higher the estimation of urgency (Jang, 2007). A relation
80 between loudness and a speaking style of "shouting" with high urgency has also
81 been reported (Mittal and Yegnanarayana, 2013; Seshadri and Yegnanarayana,
82 2009). Finally, it has been pointed out that the overall intensity of a voice plays a
83 critical factor in urgency, but this has not been studied systemically (Jang, 2007).

84 In the above studies, researchers arbitrarily manipulated the types and
85 amounts of acoustic features in synthesized speech produced by a simple text-to-
86 speech (TTS) system, in order to examine each acoustic feature's influence on the
87 magnitude of perceived urgency in speech. Therefore, it is still unclear how
88 humans actually perceive a sense of urgency in real speech and what strategies
89 they use to inform others of danger via speech. Previous studies reported that
90 prosody-related features have an important role on affective speech perception
91 (Grinkovtsova et al., 2012; Hammerschmidt and Juregens, 2007; Lieberman and
92 Michaels, 1962). For instance, by using human voice conversion techniques, Xue
93 et al. (2019) showed that the F0 and spectrum are important for emotional speech
94 perception. It has been reported that the F0 is very important for rich voice
95 expression, e.g., for artificial voices (Kawahara, Fujimura, and Konpaku, 2006;
96 Mittal and Yegnanarayana, 2015; Sudarsana and Yegnanarayana, 2019). Jang
97 (2007) reported that a decrease in speech duration to 70% increases the
98 magnitude of perceived urgency by 150%; however, it is challenging to actually
99 speak at such a fast rate. It is also hard to hear sentences spoken quickly in public
100 spaces, especially for the elderly (Konkle et al., 1977; Wingfield et al., 1985).
101 Therefore, to develop automatic evacuation announcement systems, it is

102 necessary to examine the appropriate acoustic factors for conveying danger while
103 maintaining high intelligibility.

104 Accordingly, we focus here on evacuation announcements spoken by
105 professional announcers in a real disaster. Professional speakers are trained to
106 speak in an easy-to-understand speech style (Kuwabara and Ogushi, 1984;
107 Kashimada et al., 2009; Bele, 2007; Hazlett et al., 2011). After the 2011 Great
108 Tohoku Earthquake, professional announcers on TV and radio urgently warned
109 of the need for evacuation. In addition, they consciously or unconsciously
110 changed the magnitude of perceived urgency in evacuation announcements. We
111 assume that evacuation announcements by professional announcers are both
112 intelligible and urgent, with the urgency changing gradually.

113 In this article, we describe a study that sought to examine the acoustic
114 features determining the perceived urgency of speech, in order to adjust the
115 magnitude of perceived urgency to the situation while maintaining intelligibility.
116 To focus on the speaking style, we manipulated the prosody-related features—
117 namely, the time duration, F0, and spectrum (added to the power envelope)—of
118 speech by professional announcers. We conducted a series of quantitative and
119 qualitative experiments to determine the significant acoustic features and the
120 associations among them when human participants perceived a sense of urgency
121 in real speech. Many studies have been conducted using only quantitative
122 measures to examine acoustic features' influence on the magnitude of perceived
123 urgency (e.g., Hellier et al., 2002; Jang, 2007). However, a sense of urgency is
124 considered a higher-order concept composed of multiple impressions. In our
125 previous study on auditory impressions of speech, which used the semantic
126 differential (SD) method and factor analysis (Kobayashi and Akagi, 2018), the
127 perceived urgency of speech was described with various adjectives. Thus, by
128 using psychological methods, we measured quantitative changes and qualitative
129 influences to determine acoustic features' effect on the magnitude of perceived
130 urgency.

131 In the present study, we performed two preliminary experiments and
132 three main experiments to examine the acoustic features of perceived urgency.
133 Through the preliminary experiments, we selected speech stimuli that conveyed
134 different magnitudes of perceived urgency for use in the main experiments. We

135 used the selected stimuli to synthesize new stimuli that were modulated to have
136 specific acoustic characteristics. Then, we presented these synthesized stimuli as
137 experimental stimuli in the main experiments. We conducted two experiments
138 (Experiments 1 and 2) to examine the acoustic features related to the perceived
139 urgency of speech, qualitatively or quantitatively. Then, we conducted Experiment
140 3 to examine the effects of the static and dynamic components of the F0, which
141 was the most influential acoustic feature in the first two experiments, on the
142 perceived urgency of speech. We used a magnitude estimation method in
143 Experiments 1 and 3 to enable participants to quantitatively evaluate the
144 perceived urgency of the speech stimuli, and we applied analysis of variance
145 (ANOVA) to these data. In Experiment 2, we asked participants to qualitatively
146 evaluate the perceived urgency by the SD method, and we examined these data
147 through factor analysis.

148 The rest of the paper is structured as follows. Section 2 details the original
149 speech data and the method of selecting them for speech synthesis. Section 3
150 describes the details of the signal processing used for synthesis and the acoustic
151 features that were examined in each experiment. Sections 4 and 5 describe the
152 common methodology among the experiments and the specific procedures for
153 each experiment, respectively. Finally, the results of each experiment are given
154 in section 6, and the results for the acoustic features of perceived urgency are
155 discussed in section 7.

156

157 2. Speech data

158 To experimentally examine the acoustic features related to the perceived
159 urgency of speech, we had to present speech stimuli that were manipulated to
160 highlight each feature. To this end, we used a speech synthesis tool, STRAIGHT
161 (Legacy-STRAIGHT: Kawahara, et al., 1999), to produce converted speech by
162 replacing three acoustic features in the original speech—namely, the duration, F0,
163 and spectrum—with different magnitudes of urgency. It was also necessary to
164 keep the linguistic content constant. Thus, we used five speech stimuli with the
165 same linguistic information but different speaking styles. Through two
166 preliminary experiments, these five stimuli were chosen from 57 evacuation
167 announcements spoken by professional announcers during an actual disaster. In

168 this section, we describe the preliminary experiments to select the original speech
169 stimuli and the acoustic feature replacement procedure to produce the speech
170 stimuli for the main experiments.

171

172 2.1. Evacuation speech data from actual disaster

173 We conducted two perceptual listening experiments to select speech stimuli
174 that conveyed different magnitudes of perceived urgency and could be used for
175 synthesis. These experiments were reported previously in a published work
176 (Kobayashi and Akagi, 2018), but we repeat the findings here because the
177 published paper was written in Japanese.

178 In these experiments, we used recorded speech data from actual
179 evacuation announcements when a tsunami was predicted after an earthquake
180 off the Fukushima Prefecture coast on November 22, 2016. The announcements
181 were spoken by multiple professional announcers with various speaking styles
182 on different TV channels. From the 16 hours of recorded data, we selected 57
183 evacuation announcements that had similar content (e.g., "Get away now") and
184 were spoken by 14 announcers (nine men, five women; age range: 20s–50s) with
185 different speech styles. In preliminary tests before the preliminary experiments,
186 we confirmed that participants could distinguish the speech styles of these 57
187 stimuli.

188

189 2.2. Selection of speech data with different magnitudes of perceived urgency for 190 synthesis

191 The first experiment was conducted to examine the speech style's
192 influence on the evacuation behavior. We asked 50 participants (25 men, 25
193 women; age range: 20s–60s, with 10 participants for each decade) to evaluate each
194 speech stimulus in terms of four evaluation items. Participants selected the
195 closest opinion to their own from four options provided for each evaluation item.
196 Two of the four items and their answer options were as follows: Q1: "*Would you*
197 *follow the voice instruction?*" (A1: "*yes, I would*"; A2: "*probably*"; A3: "*probably not*";
198 A4: "*no, I would not*"); Q2: "*How dangerous did you think the situation was from the*
199 *voice?*" (A1: "*very dangerous*"; A2: "*pretty dangerous*"; A3: "*somewhat dangerous*"; A4:
200 "*not dangerous*"). The percentage of participants who chose A1 was used as an

201 index of the instruction's effectiveness (Q1) and its indication of danger (Q2) for
202 each announcement. Figure 1 shows a scatter plot of the indexes for Q1 and Q2
203 for the 57 voice announcements. There was a high correlation between these
204 indexes (Pearson product-moment correlation coefficient: $r = 0.74$, $p < 0.01$),
205 which implies that the speaking styles of the evacuation announcements
206 significantly conveyed the danger and encouraged evacuation.

207 In the second experiment, we examined the auditory impressions of these 57
208 announcements by using the SD method and factor analysis. We asked the same
209 50 participants to rate each voice in terms of 16 adjective pairs by using a 7-point
210 scale. As summarized in Table 1, three factors were extracted via each adjective
211 pair's factor loading: "clarity," "urgency," and "vocal quality." As a measure of
212 each impression's strength, we calculated each factor's score for each
213 announcement; a positive score indicated greater strength. Figure 2 shows a
214 scatter plot of the "urgency" score and the Q2 index from Preliminary Experiment
215 1 for the 57 evacuation announcements. The high correlation ($r = 0.76$, $p < 0.01$)
216 suggests that the chosen announcements were appropriate for investigating the
217 magnitude of perceived urgency. In addition, we found a high correlation ($r =$
218 0.76 , $p < 0.01$) between the "clarity" and "urgency" scores, which suggests that the
219 perceived clarity of the announcements depended on the strength of the
220 impression of urgency. Furthermore, the average evaluation values for the eight
221 adjective pairs involving clarity (see Table 1) were more than 3 for all 57
222 announcements, which means that all of them received a positive rating for
223 clarity. From these results, we concluded that the clarity of the announcements
224 was preserved regardless of the speech style in terms of the perceived urgency.

225 From these preliminary results, we selected five announcements with
226 different magnitudes of urgency for conversion in the main experiments. As
227 mentioned above, several speech stimuli had the same content but different
228 prosody when spoken by the same announcer on a TV broadcast during the
229 actual disaster. Among such stimuli, the five chosen stimuli were ranked by an
230 appropriate procedure, and we thus regarded them as reasonable for examining
231 acoustic features correlated to the perceived urgency of speech. Here, we denote
232 their magnitudes as high (H), moderately high (MH), moderate (M), moderately
233 low (ML), and low (L) according to the "urgency" score. Figure 2 represents these

234 five stimuli with black circles. Hereafter, when we refer to speech stimuli, we
235 strictly mean these five stimuli. They were spoken by the same male announcer,
236 who spoke the same Japanese sentence in each case (「今すぐ逃げてください」
237 /i/ma/su/gu/ni/ge/te/ku/da/sa/i/, meaning "Get away now" in Japanese).

238 Some might consider our dataset too small for this kind of validation.
239 However, as mentioned above, our speech stimuli were taken from actual
240 broadcast announcements, and there were few announcements in which the
241 same announcer uttered exactly the same words. From these voice
242 announcements, we selected 57 speech stimuli that were sufficient to distinguish
243 different speaking styles in the preliminary tests before the preliminary
244 experiments. Accordingly, we consider these five speech stimuli to be a fair
245 dataset for this analysis.

246

247 ===== here are Figures 1, 2 & Table 1 for 1-column =====

248

249 3. Signal processing methods and acoustic features

250 The experimental speech stimuli were generated according to speech
251 parameters obtained by the STRAIGHT analysis and synthesis procedure
252 (Kawahara et al., 1999). STRAIGHT combines pitch-adaptive spectral analysis
253 with a surface reconstruction method in the time-frequency region. The
254 procedure also included F0 extraction through an instantaneous frequency
255 calculation based on the concept of "fundamentalness." Successive refinements
256 of the F0 and spectral parameter extraction procedures enabled the total system
257 to resynthesize high-quality speech. For more details, refer to the cited study
258 (Kawahara et al., 1999).

259 The time duration, F0 and spectrum of the speech stimuli were extracted
260 for synthesis. The time duration of each phoneme was measured manually. The
261 F0 and spectrum (added power spectrum) were estimated using STRAIGHT. In
262 this analysis, the FFT length was 1024 points, and the frame rate was 1 ms.

263

264 3.1. Acoustic feature replacement procedure

265 Figure 3 illustrates the procedure for replacing the F0. The time
266 information was modified to keep the speech durations of the source and target

267 stimuli the same; this had to be done before converting the F0 or the spectrum.
268 To modify the time duration, first, the speech signal was manually segmented at
269 the phoneme level for the target and source stimuli. Then, each phoneme
270 duration in the source stimulus was modified to match that in the target stimulus,
271 according to a linear ratio of their durations in each stimulus. By applying
272 STRAIGHT, an initial synthesized stimulus (source speech 2) was obtained by
273 modifying only the time duration as described above. Next, the F0s, spectrum,
274 and aperiodic component (Ap) of the source stimulus (source speech 2) were
275 extracted at 1-ms intervals by using STRAIGHT. Similarly, these features were
276 also extracted from the target stimulus. Because the time duration of source
277 speech 2 was the same as that of the target stimulus, the source stimulus' F0s
278 could be directly replaced with those of the target stimulus. The Ap and spectrum
279 of source speech 2 and the F0s of the target stimulus were then combined for
280 synthesis by STRAIGHT. Finally, the synthesized speech with F0 replacement for
281 conversion was obtained. By following this procedure, the source stimulus'
282 spectrum and Ap information were kept, but its F0s were changed to those of the
283 target stimulus.

284 Replacement of the spectrum mostly followed the same procedure as
285 for F0 replacement; the exception was the last step, in which we used the F0 and
286 Ap from the source stimulus, so that the spectrum of the target stimulus could
287 be synthesized. In this case the source stimulus' spectrum was changed to that of
288 the target stimulus, but the other information was kept. Note that the spectrum
289 here refers to the upper envelope of a speech signal's power spectrum, which is
290 related to the power of speech.

291

292 ===== here are Figures 3 for 2-column =====

293

294 3.2. Acoustic feature conditions for each experiment

295 In our main experiments described below, the source stimulus was the
296 L or H stimulus, and the other stimuli were used as the targets for synthesis. We
297 replaced the specified features of the L, ML, M, or MH target stimulus with those
298 of the H source stimulus; similarly, we replaced the specified features of the ML,
299 M, MH, or H target stimulus with those of the L source stimulus.

300 3.2.1. Experiments 1 and 2

301 The experimental speech stimuli consisted of the five original stimuli and the
302 stimuli synthesized by the above procedure. We used two types of experimental
303 conditions, which involved four magnitudes of the perceived urgency (L, ML, M,
304 MH, H) and five acoustic features. The components of the acoustic feature
305 conditions were as follows.

306 Duration (Dur): The length of the source stimulus was stretched to match that
307 of each target stimulus. Under this condition, the speech rate and pause
308 duration between phrases were also manipulated.

309 Fundamental frequency (F0s): The F0s of the source stimulus were replaced
310 with those of each target stimulus.

311 Spectrum (Spec): The spectrum of the source stimulus was replaced with that
312 of each target stimulus. Also, the RMS of the synthesized speech was
313 confirmed to be equivalent to that of the target.

314 All (All): For the source stimulus, all three of the above features were replaced
315 with those of each target stimulus. In theory, the converted speech under
316 this condition was the resynthesized target stimulus.

317 Original: The original five speech stimuli selected from Preliminary
318 Experiment 1 (L, ML, M, MH, H) were used as is.

319 Because of the synthesis process, the duration of each stimulus also varied
320 depending on the target stimulus, under all conditions. The speech stimuli were
321 presented at 58–62 dBA.

322

323 3.2.2. Experiment 3

324 As shown in Figure 4, the experimental stimuli were created by using
325 STRAIGHT to synthesize speech stimuli with the same F0 time average but
326 different F0 time variation, or with the same F0 time variation but a different F0
327 time average. First, we extracted the F0s of each stimulus by the same procedure
328 used in Experiment 1. Then, the overall time average was calculated from the F0s
329 for each stimulus. To obtain the fluctuation in F0, the frequency differences
330 between the F0s and the overall time average were calculated every 1 ms. The
331 mean F0 of the source stimulus was modified to be the same as that of the target
332 stimulus, and the stimulus was then resynthesized using the modified F0s. We

333 refer to this case as the "F0 average" condition (Figure 4A). Alternatively, the
334 mean F0 of the target stimulus was modified to be the same as that of the source
335 stimulus, the F0s of the source stimulus were replaced with those of the target
336 stimulus, and the mean F0 was modified by the same replacement procedure
337 used in Experiment 1. We refer to this case as the "F0 fluctuation" condition
338 (Figure 4B). We confirmed in advance that the mean F0 and the F0 fluctuation of
339 the synthesized speech were equal to those of the target.

340 We used the L and H stimuli as the sources for speech synthesis. For the
341 L source stimulus, the target was the ML, M, MH, or H stimulus; similarly, the
342 targets for the H source stimulus were the other four stimuli. The sound pressure
343 level was fixed at 62 dBA.

344 In addition to the acoustic feature conditions described above, we also
345 used a condition referred to as the "F0s" condition. Under this condition, both the
346 F0 averages and the F0 fluctuation were varied among those of the target stimuli
347 (Figure 4C). Accordingly, this condition was the same as the F0 component of the
348 acoustic feature conditions used in Experiments 1 and 2.

349

350 ===== here is Figure 4 for single column =====

351

352 4. Experimental methodology

353 4.1. Ethical statement

354 Informed written consent was obtained from each participant before the
355 experiments were conducted. The Japan Advanced Institute of Science and
356 Technology (JAIST) ethics committee approved all procedures.

357

358 4.2. Participants

359 A total of 48 men and women participated in the experiment as listeners.
360 In Experiment 1, 10 native Japanese speakers participated (five men, five women;
361 between 22 and 24 years old; normal hearing). In Experiment 2, 28 native
362 Japanese speakers participated (18 men, 10 women; between 22 and 24 years old;
363 normal hearing). In Experiment 3, 10 native Japanese speakers participated (five
364 men, five women; between 24 and 26 years old; normal hearing). Ten listeners

365 participated in all of the experiments. The interval between experiments 1 and 3
366 was about nine months.

367

368 4.3. Apparatus

369 The experiment was controlled using MATLAB 2015 (Mathworks) on a
370 computer (Windows 10, Intel Core i5). The stimuli were presented to both ears
371 of the listeners through a digital-to-analog (D/A) converter (RME, Fireface UCX),
372 a driver unit (STAX, SRM-1/MK-2), and electrostatic headphones (STAX, SR-
373 404).

374

375 5. Experimental procedures

376 5.1. Experiment 1: Quantitative measurement of acoustic features related to 377 perceived urgency

378 Experiment 1 was conducted to examine the influence of the acoustic features
379 of speech on the magnitude of perceived urgency through quantitative
380 psychological measurement. We converted the five original speech stimuli for
381 synthesis, as described above, by replacing the speech duration, spectrum, and
382 F0s with those of other stimuli among the five. If changing only the F0s caused
383 the synthesized speech to be evaluated as similar to the target stimulus in terms
384 of the perceived urgency, then this would mean that the F0s greatly contributed
385 to the perceived urgency. On the other hand, if this kind of change was evaluated
386 as similar to the source stimulus, then this would mean that the F0s were not
387 related much to the perceived urgency.

388 The tests were conducted in a soundproof room. The magnitude of
389 perceived urgency of each stimulus was measured using the magnitude
390 estimation method (Stevens, 1957). The listeners were asked to quantify the
391 magnitude of the urgency of each stimulus as compared to that of a control
392 stimulus with a base score of 100. For example, they were asked to give a score
393 of 200 when they perceived twice the magnitude of urgency as compared to the
394 control; similarly, they were asked to give a score of 50 when they perceived half
395 the magnitude. The original L stimulus was used as the control. Ten trials were
396 repeated for each stimulus, and the order was randomized for each listener.

397

398 5.2 Experiment 2: Qualitative measurement of acoustic features related to
399 perceived urgency

400 Experiment 2 was conducted to examine the qualitative change in perceived
401 urgency in terms of the acoustic features of speech by using the semantic
402 differential (SD) method and factor analysis, for the same speech stimuli used in
403 Experiment 1. The participants' auditory impressions of each stimulus were
404 evaluated using the SD method on a 7-point scale. As summarized in Table 2, we
405 used the same 16 adjective pairs as in Preliminary Experiment 2. The orders of
406 the speech stimuli and adjectives were randomized. The listeners were allowed
407 to listen to each stimulus repeatedly. The alignment of each adjective pair was
408 fixed for each participant in order to maintain balance of the alignments across
409 all listeners.

410

411 5.3. Experiment 3: Effects of static or dynamic component of F0 on perceived
412 urgency

413 Experiment 3 was conducted to determine the essential factor for
414 perceived urgency by manipulating the F0 time average and F0 time variation.
415 In this study, we assumed that the F0 was the sum of the F0 fluctuation and the
416 mean F0 (in terms of the log frequency). The procedure was the same as that in
417 Experiment 1.

418

419 6. Results

420

421 6.1. Experiment 1: Quantitative measurement of acoustic features related to
422 perceived urgency

423 The geometric means of the five scores were calculated for each
424 stimulus as the magnitude of perceived urgency for each listener. Figure 5 shows
425 these mean magnitudes for each converted stimulus with respect to each target
426 stimulus. Under the "All" condition, the magnitudes for the converted stimuli
427 were almost the same as those of the target stimuli. We thus conclude that the
428 synthesis process did not affect the evaluation of perceived urgency. When the
429 time duration or spectrum of the resynthesized speech was replaced with that of
430 a target stimulus, however, the magnitudes of perceived urgency differed

431 significantly. In contrast, the magnitudes of perceived urgency under the "F0s"
432 condition were almost the same as for the target stimuli.

433 We performed two-way repeated analysis of variance (ANOVA) for the
434 acoustic feature conditions (Dur, F0s, Spec, All, Original) and magnitude
435 conditions (ML, M, MH, and L or H). The results showed that the interaction
436 between the acoustic features and the magnitude was significant for both source
437 speech stimuli (L: $F_{9,159}=12.98, p < .001, \eta^2=0.06$; H: $F_{9,159}=9.94, p < .001, \eta^2=0.05$).
438 In addition, for both source stimuli, post hoc tests (Ryan's method) revealed
439 significant differences in the magnitude of perceived urgency among the "All,"
440 "Dur" ($p = 0.0001$), and "Spec" ($p = 0.0001$) conditions, and among the "F0s," "Dur"
441 ($p = 0.0001$), and "Spec" ($p = 0.0001$) conditions. However, these tests revealed no
442 significant differences between the "All" and "F0s" conditions ($p = 0.93$) and
443 between the "Dur" and "Spec" conditions ($p = 0.43$). These results indicated the
444 same tendency as in our pilot study with other listeners. Accordingly, they
445 suggest that changes in the F0s of speech have an essential effect on the
446 magnitude of perceived urgency.

447

448 ===== here is Figure 5 for single column =====

449

450 6.2. Experiment 2: Qualitative measurement of acoustic features related to
451 perceived urgency

452 For each evaluation score, factor analysis was conducted for the 16
453 adjective pairs with the main factor method and promax rotation. Because the
454 factor loading of one pair ("*stiff-soft*") was less than 0.4 for all factors, the factor
455 analysis was repeated with that pair excluded. Two factors were extracted with
456 an eigenvalue of at least 1 (see Table 2). The adjective pairs "*busy-tranquil*," "*tense-*
457 *relaxed*," and "*fervent-detached*" each had a high factor load on the first factor. We
458 considered this factor the same as "urgency" in Preliminary Experiment 2 and
459 regarded it to be related to the perceived urgency of speech. The adjective pairs
460 "*forceful-not forceful*," "*aware-unaware*," and "*well projected-poorly projected*" each
461 had a high factor load on the second factor. We considered this factor the same
462 as "clarity" in Preliminary Experiment 2 and regarded it to be related to the clarity
463 of speech.

464 Figure 6 shows the "urgency" score for each converted stimulus with
465 respect to each target stimulus. We found that this score under the "All" condition
466 was almost the same as for each target stimulus. The results were consistent with
467 those of Experiment 1, indicating that the conversion process did not affect the
468 impression of perceived urgency in speech. Moreover, the "urgency" score under
469 the "F0s" condition was close to the score for each target stimulus. In contrast, the
470 scores under the "Spec" and "Dur" conditions significantly differed from those for
471 the target stimuli. These results indicate that the F0s are essential for perceiving
472 urgency, which matches the results of Experiment 1.

473 Figure 7 shows the semantic profiles of each converted stimulus under
474 each acoustic feature condition. The scores for the "urgency" adjective pairs
475 ("*busy-tranquil*" to "*light-heavy*") varied with the F0s irrespective of the source
476 stimulus; this result indicates that the F0s affect all aspects of perceived urgency.
477 In contrast, the scores under the "Spec" and "Dur" conditions changed depending
478 on the source stimulus (L or H) irrespective of the target stimulus. The scores
479 under the "Dur" condition changed depending on the target stimulus for
480 adjective pairs related to the speech rate, such as "*busy-tranquil*" or "*fast-slow*,"
481 although they changed depending on the original stimulus for other adjective
482 pairs with a high "urgency" factor load. This result indicates that the speech rate
483 only somewhat affects the magnitude of perceived urgency. Overall, these results
484 suggest that the auditory impression of perceived urgency is strongly affected by
485 the F0s of speech.

486
487 ===== here is table 2 for 2-column =====
488 ===== here is Figure 6 for 1-column and Figure 7 for 2-column =====

489
490 6.3. Experiment 3: Effects of static or dynamic components of F0 on perceived
491 urgency

492 As in Experiment 1, we calculated the geometric means of the five scores
493 for each stimulus as the magnitudes of perceived urgency for each listener.
494 Figure 8 shows these mean magnitudes. Under the "F0s" condition, the
495 magnitude increased or decreased depending on the magnitude of the target

496 stimulus. The results supported the hypothesis that the magnitude of perceived
497 urgency is affected by the F0 information.

498 Two-way repeated ANOVA for the acoustic feature conditions ("F0
499 average," "F0 fluctuation," and "F0s") and magnitude conditions (ML, M, MH,
500 and L or H) revealed that the magnitude condition had a significant main effect
501 for both source stimuli (L: $F_{3,27} = 6.41, p < 0.01, \eta^2 = 0.16$; H: $F_{3,27} = 13.38, p < 0.01,$
502 $\eta^2 = 0.14$). For the H source stimulus, the two-way repeated ANOVA showed a
503 significant interaction between the three acoustic feature conditions and the
504 magnitude condition ($F_{6,54} = 3.21, p < 0.01, \eta^2 = 0.01$). In a post hoc test, the simple
505 main effect of the acoustic feature condition was significant for each target
506 stimulus other than MH (L: $F_{2,72} = 5.54, p < 0.01$; ML: $F_{2,72} = 9.86, p < 0.01$; M: $F_{2,72}$
507 $= 10.54, p < 0.01$; MH: $F_{2,72} = 1.94, p = 0.15$). These results ($p < 0.05$) revealed that
508 the magnitude was different among the "F0s," "F0 average," and "F0 fluctuation"
509 conditions, but for all target speech stimuli, there was no significant difference
510 between the "F0 average" and "F0 fluctuation" conditions. The post hoc test also
511 revealed that the magnitudes of perceived urgency differed between the
512 magnitude condition and all three acoustic feature conditions. For the L source
513 stimulus, there was no significant interaction ($F_{6,54} = 1.21, p = 0.31, \eta^2 = 0.02$), nor
514 was there a significant effect of the acoustic feature conditions ($F_{2,18} = 2.42, p =$
515 $0.12, \eta^2 = 0.03$). Also, the magnitude of perceived urgency under the "F0
516 fluctuation" condition was consistent with the results under the "F0s" condition
517 in Experiment 1. This agreement suggests that the results were reliable.

518 Finally, these results indicate that both the F0 time average and F0 time
519 variation influence the magnitude of perceived urgency, although each effect is
520 weak.

521

522 ===== here is Figure 8 for single column =====

523

524 7. Discussion

525 *Influences of F0 on perceived urgency*

526 This study aimed to elucidate the acoustic features determining the perceived
527 urgency of speech in order to control the urgency while maintaining clarity in an
528 automatic evacuation announcement system. By replacing the F0s and other

529 acoustic features in actual speech stimuli, we performed experiments using
530 evacuation announcements spoken by professional announcers. From the results
531 of these experiments, we argue that the F0 is the most critical acoustic feature for
532 determining the magnitude of the perceived urgency of speech. We showed
533 quantitatively and qualitatively that the F0 affects the sense of urgency. In
534 particular, the semantic profiles obtained in Experiment 2 revealed that the F0
535 affected the participants' evaluation of the adjective pair "*tense-relaxed*." In
536 contrast, the speech duration changed the evaluation of the "*busyness*" of speech
537 but had little influence on the magnitude of perceived urgency. These results
538 indicate that the impression of tension in evacuation announcements influences
539 the magnitude of perceived urgency.

540 When people announce emergencies, they feel both physical tension
541 (tightening of the muscles around the vocal cords) and mental tension. In this
542 study, we assumed that such muscle tension is reflected in the F0 information of
543 speech. Previous studies reported that the respiratory system is affected by
544 excited autonomic nervous system activities in emergencies, and glottal pressure
545 changes the F0 and sound level of speech (Scherer, 1986). We may perceive the
546 F0 as necessary for determining the magnitude of perceived urgency in speech
547 as a result of learning associations between F0 changes and the activities of the
548 autonomic nervous system in daily life. However, there is little established
549 knowledge concerning the effect of physiological arousal on speech production
550 and consequent changes in acoustic speech signals (Banse and Scherer, 1996). It
551 will thus be necessary to examine the relationships among physiological arousal,
552 speech production, and acoustic features with perceived urgency.

553 Our study results were consistent with those of previous studies
554 showing the influence of F0 modulation, but our results indicated a difference
555 concerning the dynamic components of the F0. Specifically, our study showed
556 that the magnitude of perceived urgency increases when the dynamic F0
557 fluctuation is significantly large (see Figure 7). This is inconsistent with a
558 previous study, which found that the magnitude of perceived urgency increases
559 when the F0 contour is flat rather than fluctuating (Park and Jang, 1999). Such
560 differences in F0 fluctuation, which are useful for the perceived urgency, may be
561 due to language differences. Alternatively, we assume that both flat and large

562 fluctuations are essential. The H voice has both flat parts and large-fluctuation
563 parts in the same sentence. The contrasts may be essential or may depend on the
564 part of a sentence. We consider both local and overall enhancement to occur
565 when people try to convey information desperately.

566 Because of limits in our method of feature replacement accompanied by
567 changes in the time duration, we could not strictly distinguish the effect of the
568 F0s from the effect of the time duration. We will thus need to further examine
569 prosody effects, especially the F0 fluctuation.

570

571 *Evacuation announcements in terms of affective speech and F0*

572 It is possible that our results can be regarded as a case of affective speech,
573 especially speech by a charismatic or persuasive voice, because the voices in our
574 stimuli were intended to make people obey an evacuation order. Conventional
575 research has reported that the F0 or F0 range is correlated with a voice's
576 characteristic of persuasion or leadership (Mayew et al., 2013; Neibuhr et al.,
577 2016; Signorello et al., 2020; Zoghaib, 2019). Previous studies have shown that a
578 lower pitch is positively associated with dominance or leadership (Mayew et al.,
579 2013; Zoghaib, 2019), which is inconsistent with our results. On the other hand,
580 it was recently reported that charismatic leaders convey messages by using a
581 voice with a higher overall F0, a wider F0 range, and an SPL range (Signorello et
582 al., 2020; Niebuhr, 2016); those findings are similar to our results. We cannot
583 address the differences here, but we conclude that our results were at least
584 related to the dangerous situation being conveyed in the experimental speech
585 data. The importance of the F0 in persuasive speech will thus need further
586 examination.

587

588 *Benefits of F0 modulation to control perceived urgency of speech*

589 We also point out the benefits of F0 modulation for controlling the
590 magnitude of perceived urgency in speech. In this study, we demonstrated that
591 the magnitude of perceived urgency is influenced by F0 fluctuation: a higher
592 mean F0 and larger F0 fluctuation increased the magnitude of perceived urgency.
593 Such F0 modulation does not affect the intelligibility of evacuation
594 announcements and can encourage specific evacuation behaviors. Our findings

595 also show the possibility of arbitrarily controlling the magnitude of perceived
596 urgency in evacuation announcements according to the emergency. We expect
597 that appropriate expression of urgency will lead people to recognize the danger
598 of a situation and reduce their normalcy bias during disasters. For future work,
599 we will study the relationship between temporal F0 fluctuation and the
600 magnitude of perceived urgency in more detail and develop a guideline for using
601 this relationship in synthesized speech.

602

603 *Differences in results from previous studies*

604 Finally, we describe the inconsistencies from previous findings. Our
605 study revealed that the sound pressure level and time duration are not especially
606 crucial for the perceived urgency of speech. These results differ from those of
607 previous reports (Jang, 2007). A decreased speech duration changed the
608 impression of busyness (Experiment 2) but did not change the perceived urgency
609 of speech (Experiment 1). Moreover, we found that the perceived urgency was
610 not influenced by spectrum modulation with the power envelope in Experiments
611 1 and 2. This result indicates that the effect of voice intensity is smaller than has
612 been reported. We thus suggest two possible explanations.

613 The first is that the five speech stimuli used in our study were spoken
614 by a professional TV announcer. Professional announcers are trained to speak
615 clearly and rationally at all times. The results of Preliminary Experiment 2
616 showed that these speech stimuli were heard clearly regardless of the magnitude
617 of perceived urgency. It has been reported that clear speech has specific spectrum
618 features such as formant contours or formant space (Uchasaki, 2008; Smiljanic
619 and Bradlow, 2005; Amano-Kusumoto et al., 2014; Furui, 1985) and a speech rate
620 that is not too fast. The five stimuli used for conversion in this study had
621 sufficient spectral features to sound clear, so there were no significant differences
622 between them. As a result, the replacement of spectral information had no
623 influence on the magnitude of perceived urgency. Moreover, actual speech is
624 constrained by the articulatory organs, so there was little difference in the
625 amount of each acoustic feature among the five stimuli. For example, a speech
626 rate change may be effective: it is difficult to speak clearly at 1.5 times the
627 standard speech rate.

628 The second possible explanation is that the five speech stimuli were in
629 Japanese, which distinguishes words by using pitch to accent particular mora,
630 i.e., by using F0 differences (Kawahara, 2015). Accordingly, the listeners
631 emphasized or were sensitive to pitch differences in the stimuli. In the future, we
632 will examine this influence of language in detail.

633

634 8. Conclusions

635 In this study, we examined the acoustic features controlling perceived
636 urgency in speech with clarity maintained. Through preliminary experiments,
637 we selected five speech stimuli with different magnitudes of urgency but the
638 same clarity from 57 evacuation announcements spoken by professional
639 announcers during a real disaster. In our main experiments, we used converted
640 speech stimuli that replaced three acoustic features—namely, the speech
641 duration, fundamental frequency (F0), and spectrum (voice intensity)—with the
642 features of other stimuli among the five. Both quantitatively and qualitatively,
643 the results indicated that the perceived urgency was most influenced by the F0.
644 Moreover, we examined the influences of dynamic and static F0 information on
645 the magnitude of perceived urgency. These results indicated that both types of
646 F0 information influenced the magnitude of perceived urgency in real Japanese
647 speech. Overall, our results suggest the possibility that the perceived urgency of
648 evacuation announcements is controlled by the F0, depending on the emergency,
649 when speech clarity is maintained. We will need to further examine the role of
650 F0 time variation on the perceived urgency.

651

652 Acknowledgments

653 This study was supported by the SECOM Science and Technology Foundation,
654 the JST-Mirai Program of the Japan Science and Technology Agency (Grant
655 Number: JPMJMI18D1), and the SCOPE Program of the Ministry of Internal
656 Affairs and Communications (Grant Number: 201605002).

657

658 References

659 Amano-Kusumoto, A., Hosom, J., Kain, A., Aronoff, J. M. 2014. Determining the
660 relevance of different aspects of formant contours to intelligibility. *Speech*
661 *Communication*, 59, 1–9.

662 Banse, R., Scherer, K. R. 1996. Acoustic profiles in vocal emotion expression.
663 *Journal of Personality and Social Psychology*, 70, 3, 614-636.

664 Bele, I. V. 2007. Dimensionality in voice quality. *Journal of Voice*, 21, 3, 257-272.

665 Cabinet Office Japan, Disaster Management in Japan, Tsunami Evacuation
666 Countermeasures Working Group 2012. Reports on tsunami evacuation
667 measures (in Japanese).
668 <http://www.bousai.go.jp/jishin/tsunami/hinan/pdf/report.pdf> (accessed
669 July 2, 2020).

670 Drabek, T. E. 1986. Human system responses to disaster: an inventory of
671 sociological findings. New York: Springer Verlag.

672 Edworthy, J., Loxley, S., Dennis, I. 1991. Improving auditory warning design:
673 relationship between warning sound parameters and perceived urgency.
674 *Human Factors*, 33, 2, 205–231.

675 Edworthy, J. 1994. The design and implementation of non-verbal auditory
676 warnings. *Applied Ergonomics*, 25, 4, 202–210.

677 Furui, S. 1986. On the role of spectral transition for speech perception. *Journal of*
678 *Acoustical Society of America*, 80, 4, 1016–1025.

679 Grichkovtsova, I., Morel, M., Lacheret, A. 2012. The role of voice quality and
680 prosodic contour in affective speech perception. *Speech Communication*, 54, 3,
681 414-429.

682 Hammerschmidt, K., Jurgens, U. 2007. Acoustical correlates of affective prosody.
683 *Journal of Voice*, 21, 5, 531-540.

684 Hazlett, D. E., Duffy, O. M., Moorhead, S. A. 2011. Review of the impact of voice
685 training on the vocal quality of professional voice users: implications for vocal
686 health and recommendations for further research. *Journal of Voice*, 25, 2, 181-
687 191.

688 Hellier, E., Edworthy, J., Weedon, B., Walters, K., Adams, S. 2002. The perceived
689 urgency of speech warnings: semantics versus acoustics. *Human Factors: The*
690 *Journal of the Human Factors and Ergonomics Society*, 44, 1, 1–17.

- 691 Jang, P. S. 2007. Designing acoustic and non-acoustic parameters of synthesized
692 speech warnings to control perceived urgency. *International Journal of*
693 *Industrial Ergonomics*, 37, 3, 213–223.
- 694 Kashimada, D., Ogita, K., Ishikawa, T., Hasegawa, K., Ayama, M. 2009. Effects of
695 voice training on subjective evaluation of voice quality. *The Journal of The*
696 *Institute of Image Information and Television Engineers*, 63, 12, 1818–1823 (in
697 Japanese).
- 698 Kawahara, S. 2015. The phonology of Japanese accent. In Kubozono, H. (ed.), *The*
699 *handbook of Japanese language and linguistics: phonetics and phonology*.
700 Berlin: Mouton, pp. 444–492.
- 701 Kawahara, H., Masuda-Katsuse, I., de Cheveigne, A. 1999. Restructuring speech
702 representations using pitch-adaptive time-frequency smoothing and an
703 instantaneous-frequency-based F0 extraction: possible role of a repetitive
704 structure in sounds. *Speech Communication*, 27, 3-4, 187–207.
- 705 Kawahara, H., Fujimura, O., Konpaku, Y. 2006. Voice quality of artistic
706 expression in noh: an analysis synthesis study on source-related parameters.
707 *Journal of Acoustical Society of America*, 120, 3028.
- 708 Kobayashi, M., Akagi, M. 2018. Psychological evaluation of evacuation calling
709 voice. *Journal of the Acoustical Society of Japan*, 74, 12, 633–640 (in Japanese).
- 710 Konkle, D. F., Beasley, D. S., Bess, F. H. 1977. Intelligibility of time-altered speech
711 in relation to chronological aging. *Journal of Speech and Hearing Research*, 20,
712 1, 108–115.
- 713 Kuwabara, H, Ohgushi, K. 1984. Acoustic characteristics of professional male
714 announcers' speech sounds. *Acta Acustica united with Acustica*, 55, 233–240.
- 715 Lieberman, P., Michaels, S. B. 1962. Some aspects of fundamental frequency and
716 envelope amplitudes as related to the emotional content of speech. *Journal of*
717 *the Acoustical Society of America*, 34, 7, 922–927.
- 718 Mayew, W. J., Parsons, C. A., Venkatachalam, M. 2013. Voice pitch and the labor
719 market success of male chief executive officers. *Evolution and Human*
720 *Behavior*, 34, 4, 243-248.
- 721 Mittal, V. K., Yegnanarayana, B. 2013. Effect of glottal dynamics in the production
722 of shouted speech. *Journal of Acoustical Society of America*, 130, 5, 3050-3061.

- 723 Mittal, V. K., Yegnanarayana, B. 2015. Study of characteristics of aperiodicity in
724 noh voice. *Journal of Acoustical Society of America*, 137, 3411.
- 725 Niebuhr, O., Voße, J., Brem, A. 2016. What makes a charismatic speaker? A
726 computer-based acoustic-prosodic analysis of Steve Jobs tone of voice.
727 *Computers in Human Behavior*, 64, 366-382.
- 728 Omer, H., Alon, N. 1994. The continuity principle: a united approach to disaster
729 and trauma. *American Journal of Community Psychology*, 22, 2, 273–287.
- 730 Park, K. S., Jang, P. S. 1999. Effects of synthesized voice warning parameters on
731 perceived urgency. *International Journal of Occupational Safety and*
732 *Ergonomics*, 5, 1, 73–95.
- 733 Scherer, K. R. 1986. Vocal affect expression: a review and a model for future
734 research. *Psychological Bulletin*, 99, 2, 143–165.
- 735 Seshadri, G., Yegnanarayana, B. 2009. Perceived loudness of speech based on the
736 characteristics of glottal excitation source. *Journal of Acoustical Society of*
737 *America*, 126, 4, 2061-2071.
- 738 Signorello, R., Demolin, D., Bernardoni, N. H., Gerratt, B. R., Zhang, Z., Kreiman,
739 J. 2020. Vocal fundamental frequency and sound pressure level in charismatic
740 speech: a cross-gender and –language study. *Journal of Voice*, 34, 5, 808.e1-
741 808.e13.
- 742 Smiljanić, R., Bradlow, A. R. 2005. Production and perception of clear speech in
743 Croatian and English. *Journal of Acoustical Society of America*, 118, 1677–1688.
- 744 Stevens, S. S. 1957. On the psychological law. *Psychological Review*, 64, 3, 153-
745 181.
- 746 Sudarsana, R. K., Yegnanarayana, B. 2019. Analysis of aperiodicity in artistic noh
747 singing voice using an impulse sequence representation of excitation source.
748 *Journal of Acoustical Society of America*, 146, 4446.
- 749 Uchasaki, R. M. 2008. Clear Speech. In: Pisoni, D. B., Remez, R. E. (eds.), *The*
750 *Handbook of Speech Perception*, Blackwell Publishing, pp. 207–235.
- 751 Wingfield, A., Poon, L. W., Lombardi, L., Lowe, D. 1985. Speed of processing in
752 normal aging: effects of speech rate, linguistic structure, and processing time.
753 *Journal of Gerontology*, 40, 5, 579–585.

- 754 Xue, Y., Hamada, Y., Akagi, M. 2018. Voice conversion for emotional speech:
755 rule-based synthesis with degree of emotion controllable in dimensional space.
756 *Speech Communication*, 102, 54-67.
- 757 Zoghaib, A. 2019. Persuasion voices: the effects of speaker's voice characteristics
758 and gender on consumers' responses. *Recherche et Applications en Marketing*
759 (English edition), 34, 383-110.
- 760

761 Captions:

762 Figure 1:

763 Scatter plot of the results of Preliminary Experiment 1 to examine the influence
764 of speech style on the evacuation behavior for 57 evacuation speech
765 announcements. The vertical axis indicates the percentage of participants who
766 selected "yes" for the evaluation item Q1: "*Would you follow the voice instruction?*".
767 The horizontal axis indicates the percentage who selected "*very dangerous*" for the
768 evaluation item Q2: "*How dangerous did you think the situation was from the voice?*".
769 The black circles represent the five speech stimuli used in the main experiments,
770 while the gray circles represent speech stimuli used only in the preliminary
771 experiments.

772

773 Figure 2:

774 Scatter plot of the results of Preliminary Experiments 1 and 2. The vertical axis
775 indicates the "urgency" score in Preliminary Experiment 2, which examined the
776 auditory impressions of the 57 evacuation speech announcements. The
777 horizontal axis indicates the percentage of participants who selected "*very*
778 *dangerous*" in Preliminary Experiment 1. A positive score indicates a high
779 magnitude of perceived urgency. The black circles represent the five speech
780 stimuli used in the main experiments, while the gray circles represent speech
781 stimuli used only in the preliminary experiments. The black circles are labeled
782 according to the "urgency" scores for those stimuli: high (H), moderately high
783 (MH), moderate (M), moderately low (ML), and low (L).

784

785 Figure 3:

786 Synthesis process using STRAIGHT to produce the converted speech stimuli for
787 the experiments. This figure illustrates the process for the F0s, in which the F0s
788 of the L source stimulus replaced those of the H target stimulus.

789

790 Figure 4:

791 F0 contours for an example of a stimulus used in Experiment 3, under the (A)
792 F0 average, (B) F0 fluctuation, and (C) F0s conditions.

793

794 Figure 5:
795 Mean magnitudes of perceived urgency for the converted stimuli with respect to
796 the original stimuli, across all listeners by the magnitude estimation method. The
797 gridlines show that the magnitudes of perceived urgency were the same between
798 the original and converted stimuli. The original speech used for synthesis was
799 (A) the L source stimulus or (B) the H stimulus.

800

801 Figure 6:
802 "Urgency" factor scores obtained by factor analysis and the SD method for the
803 converted stimuli. The gridlines show that the magnitudes of perceived urgency
804 were the same between the target and converted stimuli. The original speech
805 used for synthesis was (A) the L source stimulus or (B) the H stimulus.

806

807 Figure 7:
808 Semantic profiles of stimuli under the "Dur," "F0s," and "Spec" conditions.

809

810 Figure 8:
811 Mean magnitudes of perceived urgency for each converted stimulus. The
812 horizontal axis indicates the target stimulus with urgency. The original speech
813 used for synthesis was (A) the L source stimulus or (B) the H stimulus.

814

815 Table 1: Factor analysis results in Preliminary Experiment 2 with the SD
 816 method for evacuation announcements.

			1 st factor	2 nd factor	3 rd factor
			(clarity)	(urgency)	(voice quality)
1. Clarity					
Well projected	-	Poorly projected	.83	.37	.11
Powerful	-	Weak	.82	.49	-.06
Aware	-	Unaware	.81	.55	.14
Distinct	-	Vague	.80	.40	.08
Forceful	-	Not forceful	.80	.45	.08
Loud	-	Quiet	.78	.47	-.00
Pleasant	-	Unpleasant	.62	.30	-.13
Bright	-	Dark	.56	.45	.42
2. Urgency					
Fervent	-	Detached	.54	.81	.26
Tense	-	Relaxed	.44	.80	.25
Busy	-	Tranquil	.41	.80	.42
Emotional	-	Rational	.30	.75	.32
Fast	-	Slow	.44	.69	.46
3. Voice quality					
Stiff	-	Soft	.45	.46	.17
High	-	Low	.37	.43	.61
Light	-	Heavy	-.28	.09	.58
Eigenvalue			6.57	2.30	65.0
Cumulative contribution ratio			43.5	57.9	1.14

817

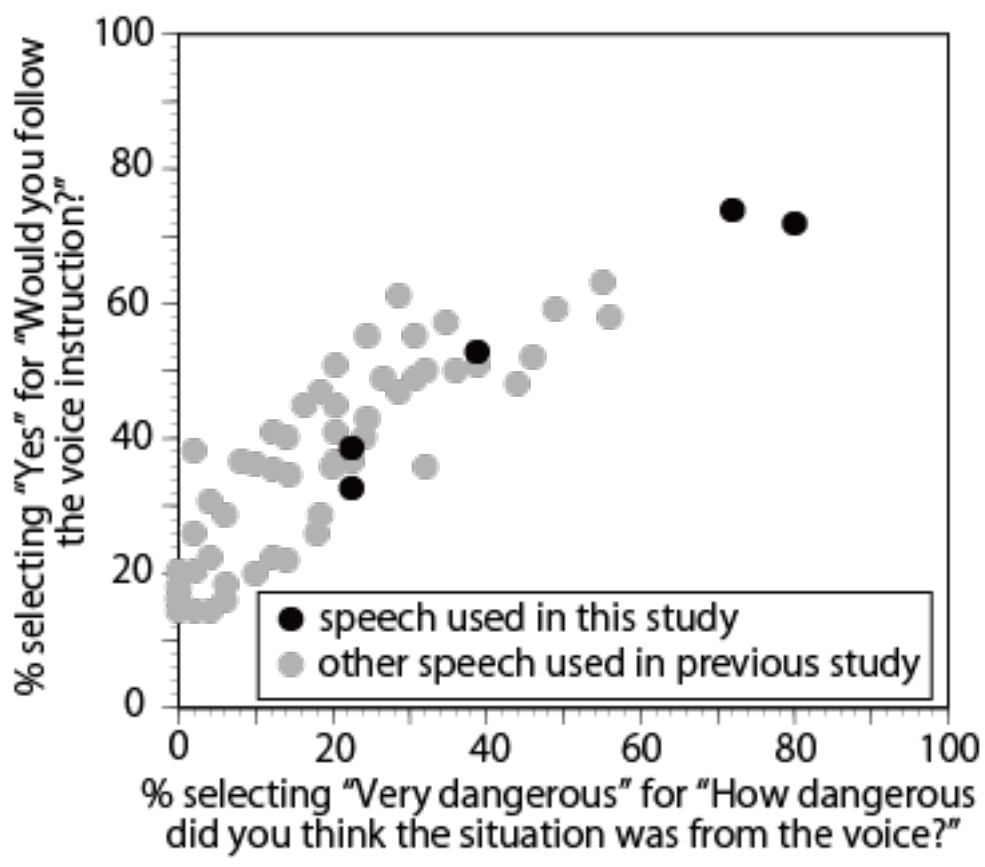
818 Table 2: Factor analysis results in Experiment 2 with the SD method for
 819 synthesized speech.

			1 st factor (urgency)	2 nd factor (clarity)
1. Urgency				
Busy	-	Tranquil	.88	-.11
Tense	-	Relaxed	.83	.07
Fervent	-	Detached	.81	.11
Emotional	-	Rational	.79	.00
Fast	-	Slow	.73	-.14
High	-	Low	.69	.04
Bright	-	Dark	.54	.31
Light	-	Heavy	.51	-.36
2. Clarity				
Forceful	-	Not forceful	.02	.77
Aware	-	Unaware	.12	.72
Pleasant	-	Unpleasant	-.19	.63
Well projected	-	Poorly projected	-.09	.62
Powerful	-	Weak	.19	.60
Distinct	-	Vague	-.10	.57
Loud	-	Quiet	.74	.53
			Eigenvalue	4.91
			Cumulative contribution ratio	32.71
				3.6
				56.7

820

821

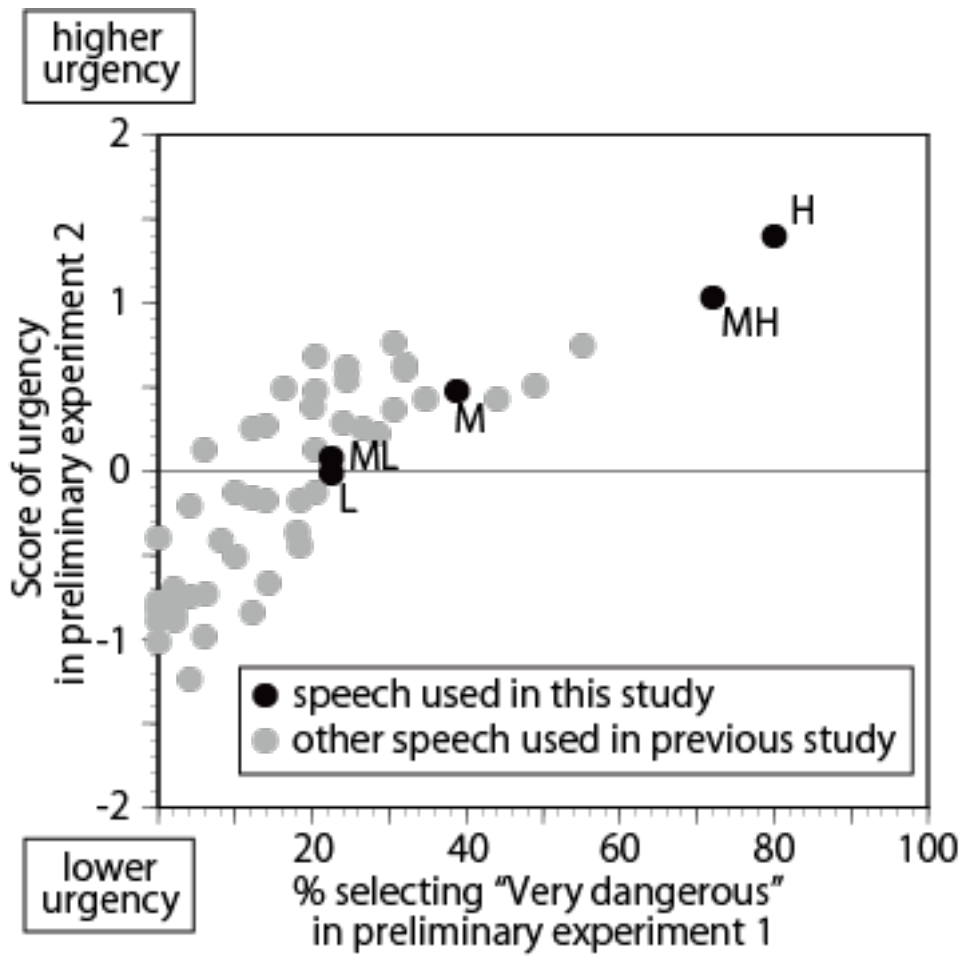
822 Figure 1:



823
824

825 Figure 2:

826



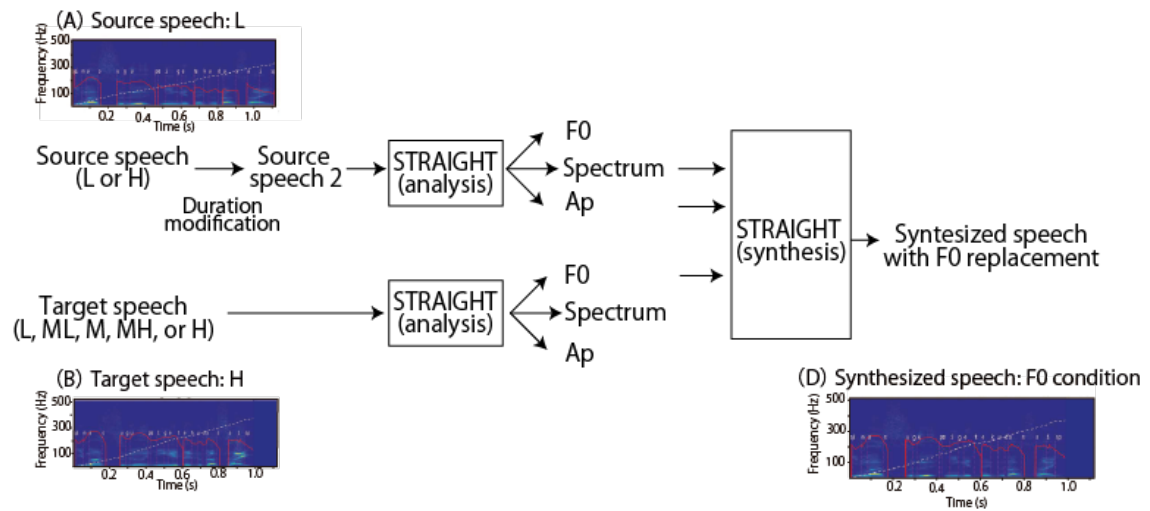
827

828

829

830 Figure 3:

831



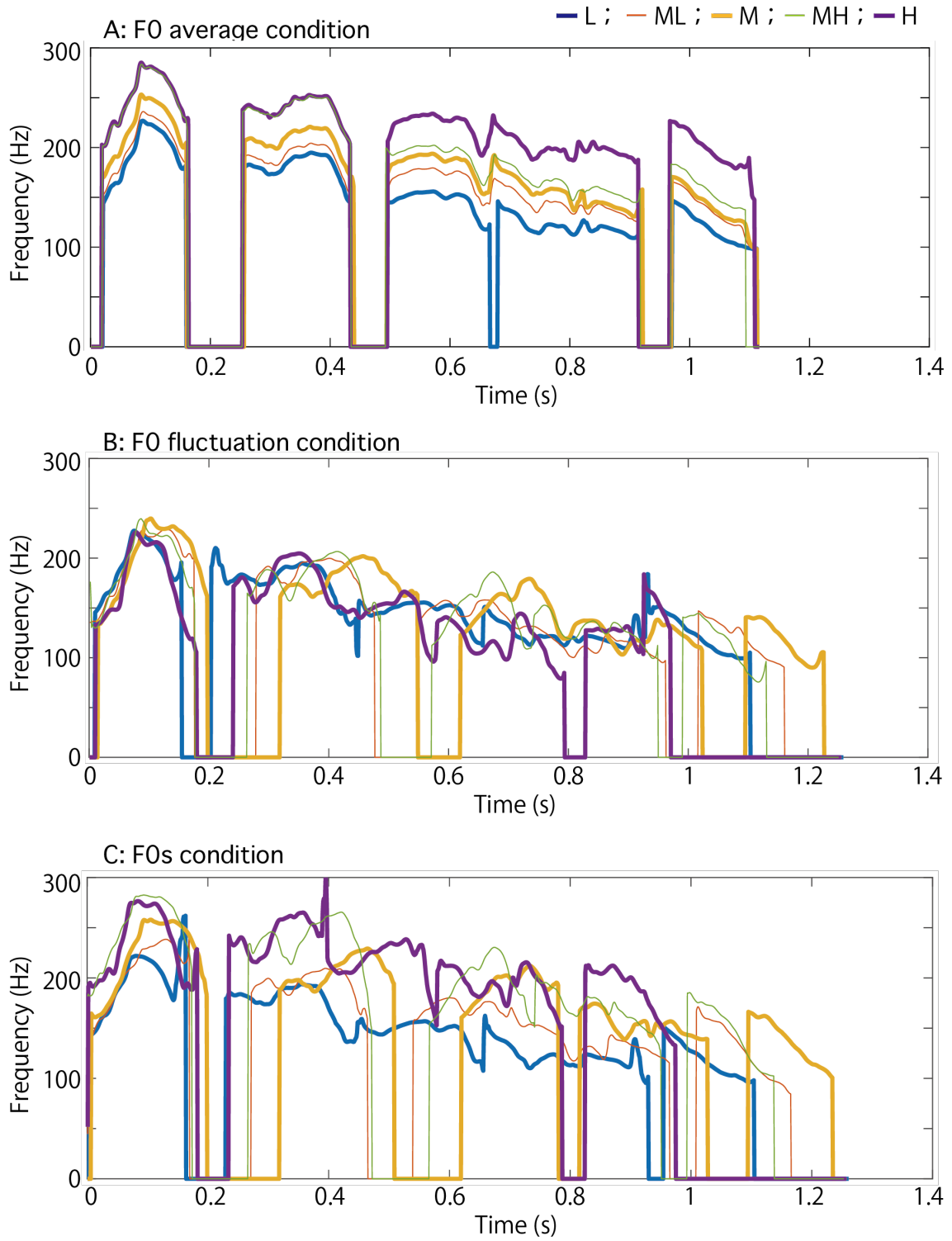
832

833

834

835 Figure 4:

836

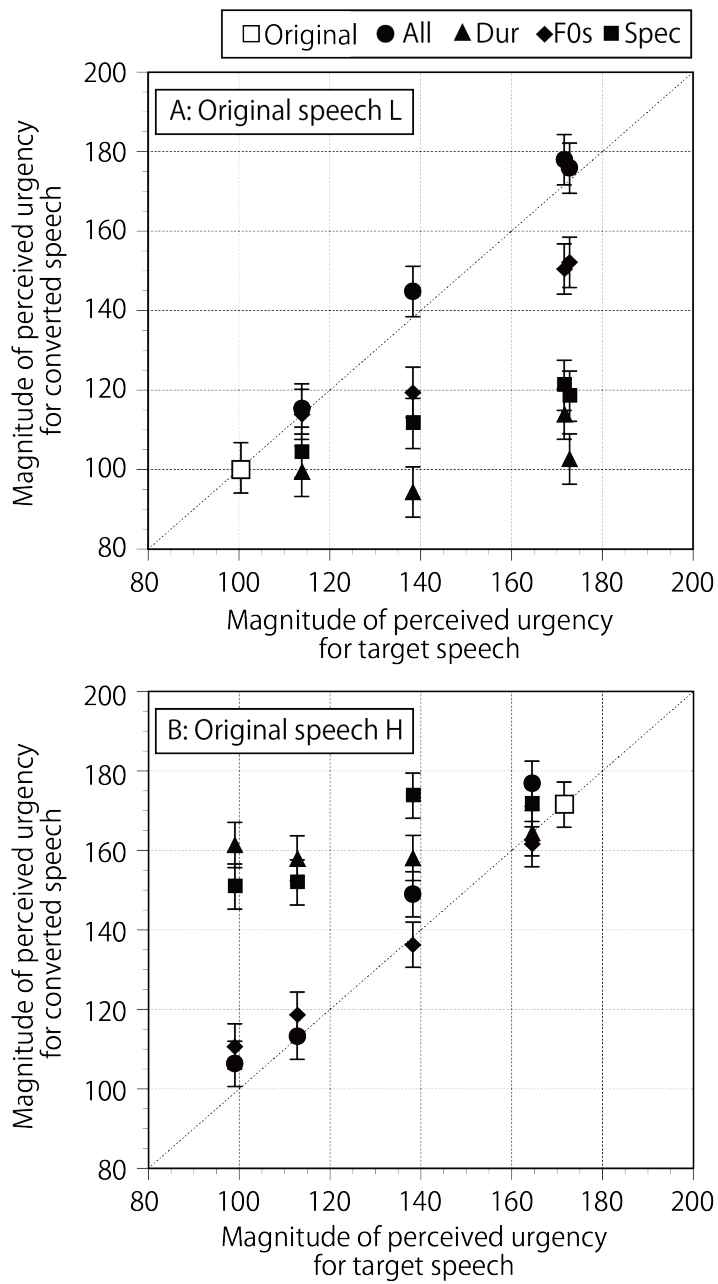


837

838

839 Figure 5:

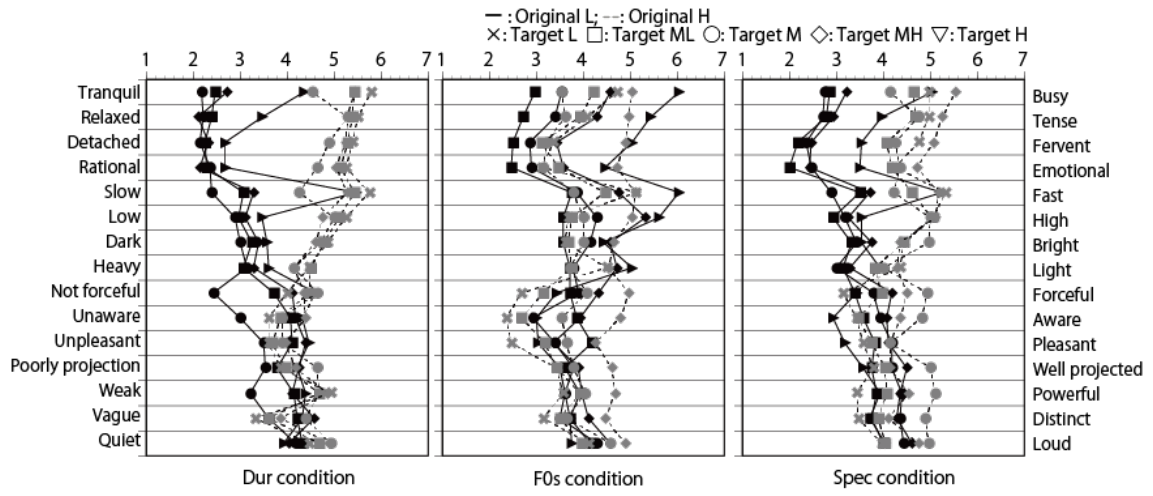
840



841

842

843



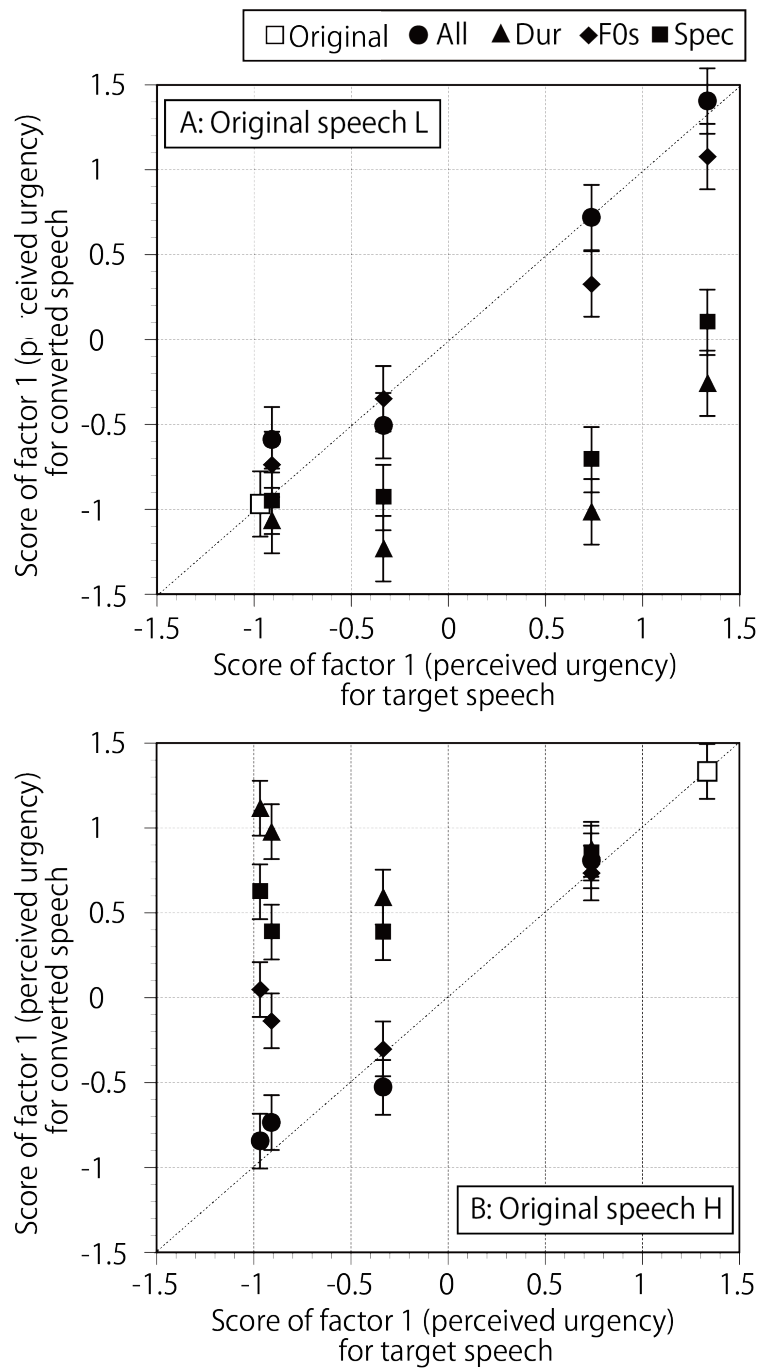
844

845

846

847 Figure 6:

848

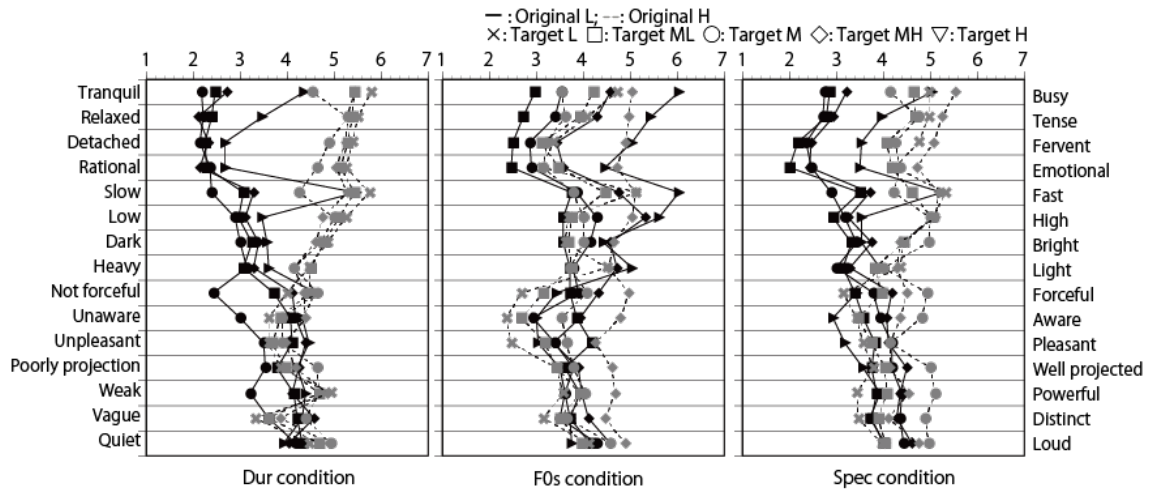


849

850

851 Figure 7:

852



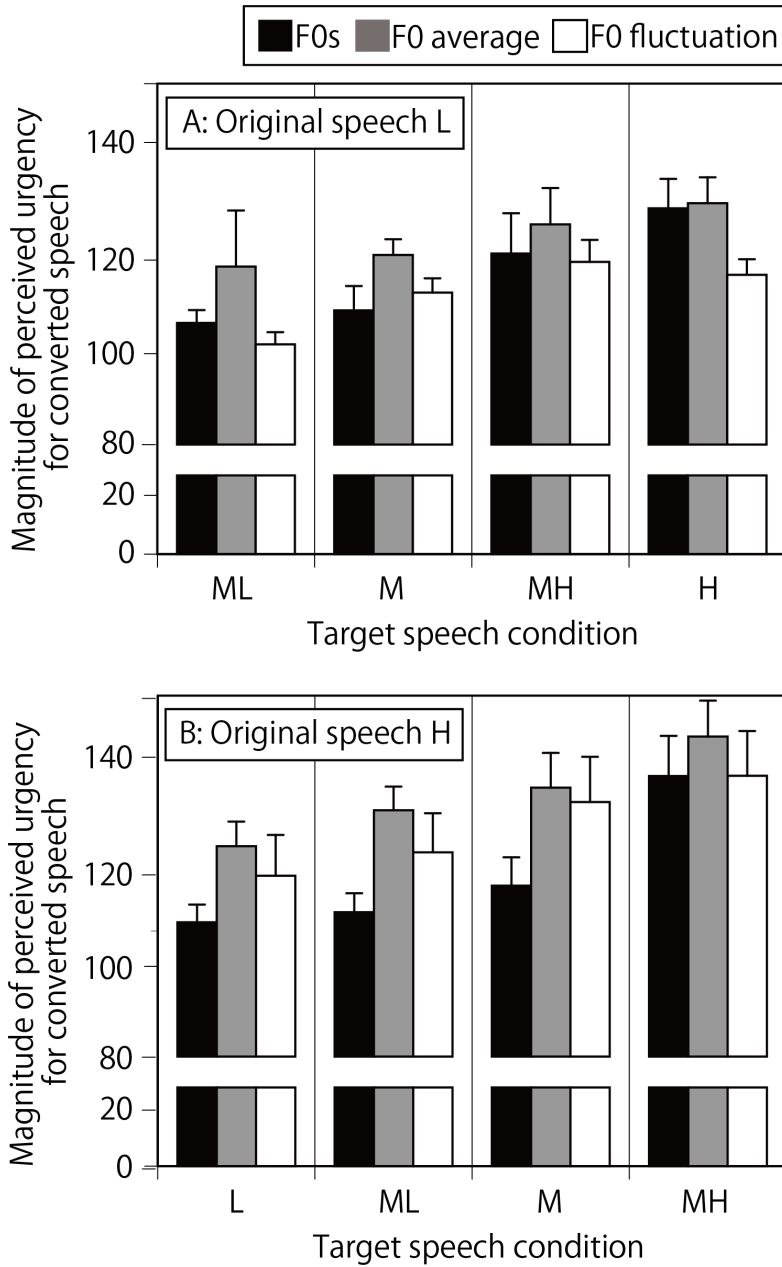
853

854

855

856 Figure 8:

857



858

859