The influence of fundamental frequency on perceived duration in spectrally comparable sounds

CAITLIN DAWSON

University of Helsinki, Department of Psychology and Logopedics, Siltavuorenpenger 1B,

00014 Helsinki, Finland

caitlin.dawson@helsinki.fi

DANIEL AALTO

University of Alberta, Rehabilitation Medicine, Communication Sciences and Disorders

and

Institute for Reconstructive Sciences in Medicine, Misericordia Community Hospital,

16940-87th Avenue, Edmonton T5R 4H5, Canada

aalto@ualberta.ca

JURAJ ŠIMKO

University of Helsinki, Department of Modern Languages, Siltavuorenpenger 1A, 00014

Helsinki, Finland

juraj.simko@helsinki.fi

MARTTI VAINIO

University of Helsinki, Department of Modern Languages, Siltavuorenpenger 1A, 00014

Helsinki, Finland

martti.vainio@helsinki.fi

July 17, 2017

Running title: Duration modulation

ABSTRACT

The perceived duration of a sound is affected by its fundamental frequency and intensity: higher sounds are judged to be longer, as are sounds with greater intensity. Since increasing intensity lengthens the perceived duration of the auditory object, and increasing the fundamental frequency increases the sound's perceived loudness (up to ca. 3 kHz), frequency modulation of duration could be potentially explained by a confounding effect where the primary cause of the modulation would be variations in intensity. Here, a series of experiments are described that were designed to disentangle the contributions of fundamental frequency, intensity, and duration to perceived loudness and duration. In two forced-choice tasks, participants judged duration and intensity differences between two sounds varying simultaneously in intensity, fundamental frequency, fundamental frequency gliding range, and duration.

The results suggest that fundamental frequency and intensity each have an impact on duration judgments, while frequency gliding range did not influence the present results. We also demonstrate that the modulation of perceived duration by sound fundamental frequency cannot be fully explained by the confounding relationship between frequency and intensity.

I.   INTRODUCTION

A simple sinusoid sound is characterized by three parameters: duration, frequency, and intensity. The subjective and objective duration of a sound (or other sensory stimulus) are not equal(1, 2) and depend on the state of the observer(3) and the organization of consecutive sounds in time(4), as well as the structure of the sound signal itself. In particular, sound signals with larger amplitudes are perceived as longer when compared with sounds of the same objective duration but smaller amplitude, indicating that sound intensity contributes to duration perception(5). The perceived duration of sounds is also prolonged by frequency, with higher tones being perceived as longer than lower ones; this has been documented for simple sinusoid sounds (6) as well as for more spectrally complex tones (7, 8, 9). Analogous modulation phenomena exist in visual and tactile sensory modalities (11, 12, 13, 14).

However, it is not clear whether intensity and fundamental frequency can influence the perceived duration of sounds independently, and to what extent. Since the frequency of sinusoidal sounds contributes to their intensity as well as to their perceived loudness (10), the modulation effect of frequency on perceived duration could potentially be explained as a byproduct of the modulation of perceived duration by intensity alone.

The duration coding mechanism that might be partly responsible for the observed frequency modulations in duration perception could be based on two different phenomena: first, the neural spikes corresponding to the onset or offset of the sound may be delayed as a function of intensity and phase-locked modulation frequency (i.e. fundamental frequency) (9) (see also (33) for neurophysiological evidence); second, the accumulation of input spikes may be affected by the intensity and frequency of the sound (as in the dual klepsydra model (34)).

In addition to low level mechanisms, several high level explanations for the duration modulation by frequency have been proposed. According to a Gestalt approach, a high frequency sound is perceived as intrinsically smaller – and consequently, shorter – than a low frequency

43 sound (36). Also, speech-like auditory objects may be subject to a compensatory strategy

44 as high pitched syllables in many languages (e.g., lexical tones in Mandarin Chinese) are

45 produced shorter than low-pitch tones (28). (This explanation can, of course, be reversed,

46 with pitch-dependency of the production patterns interpreted as a compensatory consequence

47 of perceptually driven pitch-modulation of perceived duration, e.g., (44)).

48 Disentangling influences of intensity and fundamental frequency on perceived duration of

49 a sound is demanding for several reasons. First and foremost is the fact that intensity and

50 frequency are themselves not independent from each other. Physically, intensity of a (simple)

51 sound is a function of its frequency. As equal loudness contours show, the sound frequency

52 also positively contributes to its perceived loudness up to about 3.2 kHz.

53 Apart from purely psychoacoustic considerations, the interactions between fundamental

54 frequency, intensity and duration perception are also of linguistic and phonetic interest.

55 Relative prominence of speech constituents, encoding lexical stress, emphasis, etc., is signalled

56 through a complex, context-sensitive and language-dependent interplay between these three

57 sound characteristics (as well as spectral properties), which underline the necessity of stimulus

58 choice in experimental settings. While sinusoidal sounds are an important research tool, they

59 are scarce in biologically plausible sound environment of human communication where sounds

60 are complex and the fundamental frequency is constantly moving. In tone languages, syllable

61 duration cannot be separated from the underlying tone (23, 24, 28). In quantity languages, the

62 fundamental frequency (f_o) contours contribute to distinguishing between quantity categories

63 (25, 26, 27). In addition to f_o level, its movement also interacts with perception of syllable

64 duration, albeit in somewhat language-dependent way (7, 8, 22, 28, 29, 30); this interaction

65 may depend on the linguistic or non-linguistic goals of the experimental paradigm.

66 In the present work, we revisit this issue with a special focus on disentangling the contributions

67 of fundamental frequency and intensity to the perceived duration of a sound. We report the

68 results of two forced choice experiments in which the listeners were asked to compare the

durations and intensity levels, respectively, of two sounds varying simultaneously in multiple dimensions of duration, fundamental frequency and intensity. The possible confounds between fundamental frequency and intensity are addressed in two ways. First, the sounds presented to the participants have rich spectral content that is band pass filtered to a narrow spectral frequency range well separated from the fundamental frequency. This results in a pitch sensation that corresponds to the fundamental frequency although almost no energy is left at that frequency. This percept is sometimes named the missing fundamental phenomenon since the auditory nerve fibres corresponding to the fundamental are not encoding the signal. Instead, the amplitude modulated movement of the basilar membrane at the band pass frequencies conveys the information with slow oscillations corresponding to the fundamental. Second, using logistic linear regression analysis, the effects of fundamental frequency on both duration and intensity judgments are quantified. Comparing these influences allows us to evaluate whether the contribution of fundamental frequency to loudness is sufficient to explain its effect on duration perception.

II. METHODS

A. Participants

In a pilot phase and in an earlier work where similar stimuli were used, the fundamental frequency had an impact on duration judgments of each participant (30). Since every individual recruited for the study was expected to show fundamental frequency modulation of the duration judgment on an individual level, there was no constraint to the minimum sample size. If a participant was not affected by fundamental frequency in their duration judgments, they would be excluded from the analyses since the experimental question for these individuals would not make sense.

Eleven native monolingual Finnish speakers aged 18-40 participated in the experiment. They

93 were screened for normal hearing (<20 dB). The stimuli were presented through headphones

94 that were calibrated for each participant so that the standard sound always had a fixed

95 intensity level of 66 dB SPL (A-weighted) measured by a PeakTech 5055 sound level meter.

96 Every participant took part in both discrimination tasks reported here[1]. The experiments

97 were performed according to the guidelines of the Declaration of Helsinki at the University of

98 Helsinki, Finland; the Committee for ethical review granted ethical approval to carry out the

99 study within its facilities and the participants gave their written consent to take part in the

100 experiments.

B. Procedure

102 Two 2-alternative forced choice discrimination tests were performed to evaluate multifeature

103 intensity and duration discrimination, respectively. The sounds presented to the participants

104 were varied in four parameters: duration, f_o level, dynamic f_o range, and intensity level,

105 and were drawn at random so that the probability distribution of each parameter formed a

106 truncated normal distribution.

107 The tasks were performed in two blocks (intensity discrimination block and duration discrim-

108 ination block) and the order of the blocks was randomized between the subjects. In each

109 block, the participants were presented with 300 pairs of stimuli. After presentation of each

110 pair they were asked to identify which of the two sounds was louder or longer, respectively,

111 by typing a (for the first) or x (for the second) on a standard (Finnish) QWERTY keyboard.

---

[1]In addition, the participants also performed two additional experiments preceding the discrimination tasks reported here.

C.   Signal generation

Every sound stimulus was fully characterized by four parameters—duration, f_o level, dynamic f_o interval and intensity level—drawn for each sound from a truncated normal probability distribution (if a randomly generated parameter was more than two standard deviations away from the mean, it was discarded and a new value was generated). The duration of each sound was drawn independently from the normal distribution with mean 300 ms and standard deviation 75 ms. f_o level followed a truncated normal distribution with mean 150 Hz (corresponding to 0 semitones) and a standard deviation of 4 semitones (hence, there were more sounds over 200 Hz than under 100 Hz). Here the semitone is defined as the twelfth root of two, i.e., the fixed ratio between frequencies of adjacent keys of an equally tempered keyboard. The dynamic f_o range followed a truncated normal distribution with mean 0 (static sound) and standard deviation of 4 semitones, and the intensity level followed a truncated normal distribution with mean 66 dB and standard deviation 3 dB. Altogether, the durations varied from 150 ms to 450 ms and the instantaneous pitch of the sound varied between 75 Hz and 300 Hz (this is 8+4 semitones below and above 150 Hz). Since the gammatone filter used for signal generation (see below) had a center frequency over 3 kHz, even the highest instantaneous f_o had only unresolved harmonics within the band. At the biomechanical level of the cochlea, the basilar membrane vibrations are amplitude modulated to the fundamental frequency, creating the same pattern in the auditory nerve signal. With sufficient separation (as here) between the active spectral band and the fundamental frequency, the individual harmonics of the sound should not be discernible. The intensity levels varied between 60 dB and 72 dB. Because the distributions are truncated, the true standard deviations of the generated distributions are slightly smaller.

For each randomly generated set of parameters, a positive sawtooth wave of the given duration was frequency modulated so that the instantaneous fundamental frequency was exponentially increasing/decreasing depending on the sign of the interval, and the frequency at the middle

138 of the signal duration corresponded to the given f_o level (see Fig. 1, A). To avoid a jump in

139 the waveform at the end of the signal, the predetermined duration was prolonged to the next

140 zero crossing. The sawtooth waveform was selected to ensure a rich spectral content on the

141 targeted frequency band.

142 The sawtooth wave was subsequently band-pass filtered using a 4th order approximation

143 of the gammatone filter with center frequency $f_c =$ 3141.6 Hz (39) (Fig. 1, B). The center

144 frequency was chosen so that the fundamental frequency was always well separated from the

145 frequency band containing the signal energy. Further, a normalization of the energy of the

146 signal was performed to even out residual loudness differences; the normalization was based

147 on the energy average over the first 100 ms of the waveform. Finally, the signal amplitude

148 was changed to reach the desired intensity level, completing the single sound generation.

149 Having generated the pair of stimuli for each trial in this way, the sounds were joined using

150 randomly generated inter-stimulus interval: a duration was drawn from a truncated normal

151 distribution with mean 800 ms and standard deviation 10 ms, and used as an interval from

152 the onsets of the first to the onset of the second sound. Then, 600 ms of silence was added

153 before the onset of the first sound and after the offset of the second sound and a white

154 noise was added (to the entire signal, including these pre- and post-stimulus intervals) with

155 10 dB signal to noise ratio in order to mask nonlinear distortion tones and guarantee a

156 narrow spectral band. In the pilot phase, the stimuli presented without noise did not have a

157 perceivable distortion product but for very similar sounds presence or absence of noise has

158 reportedly changed their pitch percept (45). Finally, linear onset and offset ramps covering

159 the first and the last 200 ms of the entire stimulus containing the two sounds were created.

160 Fig. 2 shows an example of one trial.
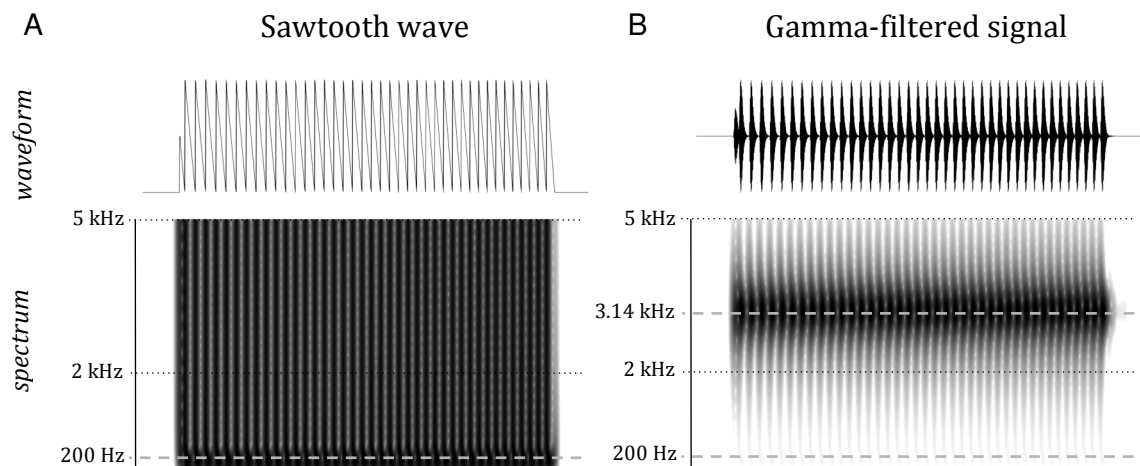
Figure 1: *The waveform and spectrum of the original (A) and gamma-filtered (B) sawtooth wave. The wave mid-point fundamental frequency is 200 Hz, duration 200 ms. The frequency movement interval is 4 semitones. Note that the energy in the filtered signal is centered around the gammatone filter's center frequency of 3.14 kH, and is negligible around the signal's fundamental frequency of 200 Hz.*
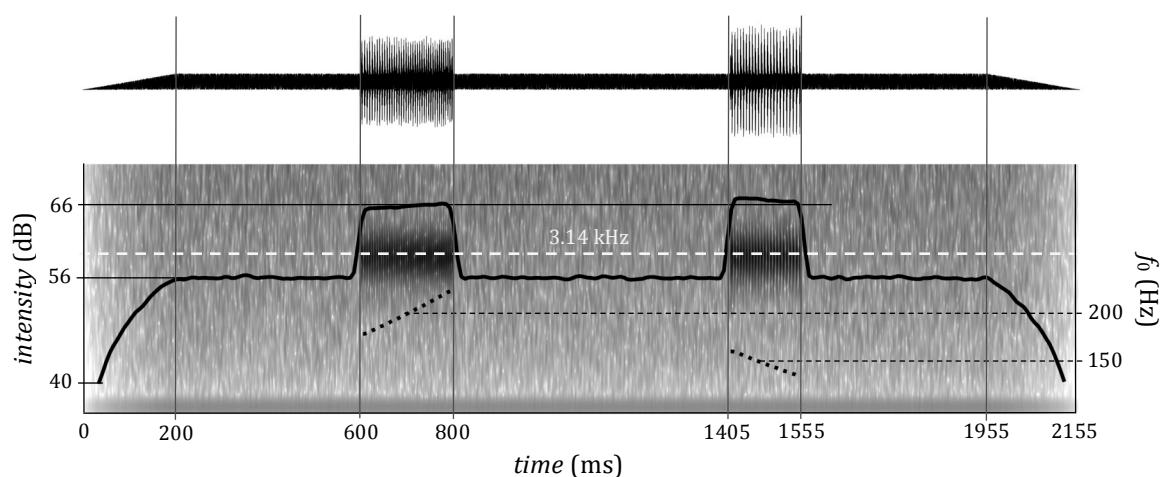
Figure 2: *An example of a trial with two stimuli with waveform (top) and spectrum (bottom). The fundamental frequency and duration parameters of the first stimulus are identical with that in Fig. 1, the second stimulus is 150 ms long, with central frequency of 150 Hz and f_o movement interval -3 semitones. The gain for the first signal is 0 dB, for the second 0.5 dB. Inter-stimulus interval of this example trial (between the onsets of the sounds) is 805 ms. The central frequency of the gammatone filter is superimposed over the spectrogram.*

D. Predictability and correlations between the design factors

Because the final periods of the sounds were completed, duration and fundamental frequency were somewhat correlated. In order to evaluate the amount of resulting interaction between the variables and its impact on the actual durations of the stimuli, linear regression models were fitted to the actual acoustic characteristics of the stimuli that were played to the participants.

A third order non-linear regression model with the sound duration as a dependent variable and f_o level and dynamic f_o interval as predictors explained 0.02% of the durational variation with a significant effect of the f_o level; the dynamic f_o interval effect was not significant. The small artificial lengthening of the sounds thus does not provide strong cues to the stimulus duration. In fact, the amount of durational manipulation was typically small, on average 1.5 ms.

The f_o gliding speed, on the other hand, gives much more information. A linear model with gliding speed as an independent and duration as a dependent variable fitted to the generated stimuli explained 9.5% of the variation and a 15th order nonlinear model explained 11.6% of the variation (increasing the order of the model beyond 15 did not increase the r^2 value any more). The effect size as estimated from the (first order) linear model is quite large: the statistical model predicts a 23 ms shorter duration for a sound with a change rate of 10 semitones per second as opposed to a static sound of the same duration. This is the largest single source of bias in the data; however, it should not affect the results since participants would not know that the gliding speed is related to duration and they should not be able to detect it during the experiment.

In fact, the design of the current work was such that these biases work, in general, against the hypotheses that higher and more dynamic f_o are perceived as longer. As the last period of each generated sound was prolonged up to the end of the last period, the low f_o stimuli and falling sounds were effectively lengthened more than the high/rising ones. Additionally,

187 in the current design, the probability of a long sound was reduced by observing a very fast

188 moving stimulus—if a subject inferred this relationship, she or he would consequently judge

189 the more dynamical stimuli as shorter rather than longer.

E. Statistical analysis of the response data

191 Separate mixed effect logistic regression (logit) models were fitted for each discrimination

192 task, with participants' response as a dependent variable (with values 1: the first sound

193 louder/longer, and 0: the second sound louder/longer). The fixed effects were the differences

194 between the actual signal parameters of the two stimuli in each trial (a value for the first

195 sound minus the corresponding value for the second sound): the difference between durations

196 (in s), the difference between intensity levels (in dB), the difference between f_o levels (in

197 semitones), and the difference between the dynamic f_o ranges.

198 For the last difference variable we tested two alternative versions, the first representing the

199 difference in absolute dynamicity (the difference between *absolute* values of the f_o ranges), and

200 the second representing the difference in dynamicity including direction of the fundamental

201 frequency slope (the difference between the *raw* values of the f_o ranges). All the manipulated

202 variables (differences) were used as random slopes for the subjects.

203 Full models with interactions among these fixed effects as well as models without interactions

204 were fitted. Statistical significance was estimated using Monte Carlo simulations as imple-

205 mented in the R package lme4 (41). Deviance reduction was calculated for the individual

206 models in order to assess the quality of fits (42).

207 In addition, in order to compare the performance of individual participants with the group

208 effects (captured by the mixed effect models), we fitted simple logistic models for the same

209 dependent and fixed effect variables as above (without interactions) for each participant

210 separately.

III. RESULTS

A. Intensity discrimination

Logistic regression analysis of the data shows that duration difference, f_o difference, and intensity difference all had a significant impact on intensity judgments. The dynamic f_o range difference did not reach significance, whether conceived as a difference in absolute (Table 1) or raw (Table 2) values of the range. The interactions were not significant, therefore only the coefficients from the models containing just the main effects are reported.

Table 1: Mixed effects model fitted to the responses of *intensity discrimination* with frequency range difference calculated as the difference between the **absolute values** of the dynamic f_o ranges.

Effect	size	error	z value	p (MCMC)
Intercept	0.14	0.076	1.8	0.07
Duration difference	6.5	0.71	9.2	$2 \cdot 10^{-16}$
Intensity difference	0.34	0.055	6.2	$4 \cdot 10^{-10}$
Frequency difference	0.14	0.036	3.9	$1 \cdot 10^{-4}$
Frequency range difference	0.028	0.020	1.4	0.2

Table 2: Mixed effects model fitted to the responses of *intensity discrimination* with frequency range difference calculated as the difference between the **raw values** of the dynamic f_o ranges.

Effect	size	error	z value	p (MCMC)
Intercept	0.14	0.0768	1.8	0.07
Duration difference	6.6	0.68	9.6	$2 \cdot 10^{-16}$
Intensity difference	0.34	0.054	6.3	$2 \cdot 10^{-10}$
Frequency difference	0.14	0.036	3.9	$1 \cdot 10^{-4}$
Raw frequency range difference	-0.006	0.011	-0.57	0.56

The effects of duration, intensity, and fundamental frequency differences are all positive and highly significant, indicating that increases in any of these parameters are associated with a perceived increase in loudness (the greater the difference, the relatively greater the parameter value for the first sound, and, according to the model, the greater likelihood of a 'first sound

222   louder' response). Also relevant are the almost identical values of the estimates of these fixed

223   effects between the two models with different dynamic range comparison variables.

224   By comparing deviance of the experimental models to the deviance of the null model which

225   includes only random intercepts for the subjects, the current models both reduce approximately

226   26 % of the deviance for intensity judgments, showing that the addition of the modulating

227   variables produces a better fit to the data for both models.

228   Comparing the model parameters allows us to compare the relative effects of sound manipu-

229   lation between different acoustic parameters. Increasing the intensity level of the first sound

230   by 1 dB adds 0.34 to the linear combination of the dependent variables in the inverse logit

231   function fitted by the model. The same quantitative effect can be achieved by 52 ms (0.052 s)

232   durational lengthening of the sound ($6.5 \times 0.052 \simeq 6.6 \times 0.052 \simeq 0.34$) or 2.4 semitones

233   increase in fundamental frequency ($0.14 \times 2.4 \simeq 0.34$).

234   The standard deviations of the random slopes were 1.8 and 1.6 for duration, 0.18 and 0.17

235   for intensity, 0.11 and 0.11 for fundamental frequency, and 0.05 and 0.02 for frequency range

236   (respectively, for the two models). Comparing these estimates to the effect sizes indicate more

237   inter-subject variability in frequency range response compared to the other signal parameters

238   (for this variable only, the standard deviation is actually greater than the estimate).

239   This assessment is explicitly confirmed by the separate logistic models with the same dependent

240   and independent variables fitted for each participant individually (the estimates and their

241   significances are reported in full in the Appendix Tables 1 and 2). The duration difference and

242   intensity difference variables are significant and positive for both versions of the model and

243   for *all* eleven participants. Also, the effect of f_o difference is significantly positive for all but

244   two subjects (subject number 1 and 8) in both models. On the other hand, the dynamicity

245   difference is only significant (and positive) for one participant in the case of difference in

246   absolute range and two participants (one negative, the other positive) for the directional

247   difference in raw dynamic ranges.

B.   Duration discrimination

Duration difference, intensity difference, and f_o difference had a significant effect on duration discrimination, but dynamic f_o range difference did not reach significance (Table 3). The interactions were not significant; therefore only the coefficients from the models containing just the main effects are reported.

Table 3: Mixed effects model fitted to the responses of *duration discrimination* with frequency range difference calculated as the difference between the **absolute values** of the dynamic f_o ranges.

Effect	size	error	z value	p (MCMC)
Intercept	0.47	0.19	2.4	0.016
Duration difference	29	2.8	10	$2 \cdot 10^{-16}$
Intensity difference	0.073	0.018	4.1	$4 \cdot 10^{-5}$
Frequency difference	0.19	0.029	6.8	$1 \cdot 10^{-11}$
Frequency range difference	0.021	0.020	1.0	0.3

Table 4: Mixed effects model fitted to the responses of *duration discrimination* with frequency range difference calculated as the difference between the **raw values** of the dynamic f_o ranges.

Effect	size	error	z value	p (MCMC)
Intercept	0.45	0.19	2.4	0.018
Duration difference	29	2.9	10	$2 \cdot 10^{-16}$
Intensity difference	0.072	0.018	4.0	$8 \cdot 10^{-5}$
Frequency difference	0.19	0.029	6.9	$5 \cdot 10^{-12}$
Raw frequency range difference	0.035	0.014	2.5	0.014

Comparing deviance of the experimental models to the deviance of the null model which includes only random intercepts for the subjects, both current model reduce approximately 45% of the deviance for duration judgments, showing once again that the addition of the modulating variables produces a better fit to the data.

Using the same technique as in the previous section shows that the duration judgments were equally impacted by 10 ms duration increase, an intensity increase by 4.0 dB, and a 1.5 semitone fundamental frequency raise. An intensity increase by 1 dB corresponds to

260 fundamental frequency increase of 0.38 semitone.

261 These findings hold for both models with the alternative definitions of the effect of dynamic

262 f_o range effects. However, for duration discrimination task, the directional effect of the f_o

263 dynamicity reached significance (Table 4) while the effect of difference in the absolute values

264 of dynamic ranges did not (Table 3). The significantly positive sign of the raw frequency

265 range difference estimates in Table 4 means that the stimuli with the rising f_o (positive

266 range value) tended to be judged as longer compared with the falling (or less rising) stimuli.

267 (Reporting an equivalence between the range difference and the other difference variables,

268 although straightforward to derive, would be very cumbersome and is therefore left out from

269 this description).

270 The standard deviations of the random slopes were 8.5 and 8.8 for duration, 0.09 and 0.08 for

271 frequency, 0.04 and 0.04 for intensity, and 0.03 and 0.03 for frequency range (respectively, for

272 the two models). That is considerably smaller than the effect size for duration, and about half

273 the estimate for frequency and intensity differences. This measure of inter-speaker variability

274 was greater than estimate for the model using difference of the absolute range values and

275 approximately the same as the estimate in the alternative model reported in Table 4.

276 The logistic regression models fitted for individual participants (see Appendix Tables 3 and 4)

277 shed further light on the degree of robustness of effects regarding the participants' judgments.

278 For both duration and f_o level differences, the effect is consistently positively significant for

279 all subjects. Somewhat surprisingly, the intensity difference effect is only significant for three

280 participants out of 11 (subject numbers 5, 7 and 11 for both model types); however, the

281 sign of the effect is also positive for all remaining participants even though the effect fails to

282 reach significance. A similar situation arose for the model with raw range difference as the

283 fixed effect (Appendix Table 4): although the estimates are positively significant for only two

284 participants (number 3 and 11); in all other cases (except participants 1 and 2) the effect

285 is also positive. On the other hand, for the model with difference in absolute range values

286 (Appendix Table 3), the sign of the estimate is positive and negative for approximately half
287 of the participants each, indicating a much greater degree of qualitative variability across our
288 participants.

289 Furthermore, in a hierarchical series of models for duration judgment, a model including the
290 intensity (but not frequency) term reduced 37.6% of the model deviance, whereas the model
291 with the frequency (but not intensity) term reduced as much as 44.5% of the model deviance
292 compared to 45.2% for the full model. This suggests that the impact of fundamental frequency
293 on duration judgment is not entirely explained by the impact of intensity. Additionally, when
294 interaction terms were added to the models, they were small and not significant.

295 C. Comparing the discrimination results

296 The two tasks were based on identically generated stimuli allowing for the models for duration
297 and intensity discrimination to be directly compared. While the leading term in both models
298 corresponds to the primary acoustical correlate of perceived duration and loudness respectively,
299 the duration and loudness judgments were both influenced by other sound parameters. Some
300 (or all) of the fundamental frequency variation could lead to variations in perceived loudness
301 that would then lead to duration modulation.

302 To interpret the results, a thought experiment is carried out where the impact of loudness, as
303 opposed to the intensity, on duration judgments is estimated from the two experiments. As-
304 suming that the impact of loudness (perceived intensity) on perceived duration is proportional
305 to the measured impact of intensity (since this is the leading term in the logistic regression
306 model for loudness judgments), and assuming that a portion of fundamental frequency would
307 impact the loudness which would then in turn impact the duration, the two logistic models
308 can be combined by substituting the intensity term in the duration judgment model with
309 a loudness term, which is a linear combination of duration, fundamental frequency, and
310 intensity.

311 The quantitative estimate of fundamental frequency induced loudness effect does not support
312 the idea that duration modulation by fundamental frequency would be primarily generated
313 by intensity variations. In the loudness judgment phase, there was an equal impact of 1 dB
314 and 2.4 semitones raise, whereas in the duration judgment phase, there was an equal impact
315 of 1 dB and 0.38 semitones raise. To give an upper limit for fundamental frequency impact
316 on the duration judgment through the loudness, not more than 0.38 semitones could be
317 bundled to the loudness percept corresponding to 15.8 % (0.38/2.4) of the total fundamental
318 frequency effect.

319 **IV. DISCUSSION**

320 Clearly, duration perception is closely linked to both intensity and fundamental frequency.
321 Higher fundamental frequency and greater intensity have been shown before to be associated
322 with longer perceived duration. The presented set of experiments targeted the question of
323 whether the impact of fundamental frequency on duration judgments is simply a confound,
324 i.e., whether it can be explained through the influence of frequency on perceived intensity.

325 The results strongly suggest that this is not the case, and that fundamental frequency
326 has an independent contribution to the perceived duration of a sound. In the light of the
327 design presented here, this finding is supported on several levels. First, care was paid to
328 neutralizing the interdependencies between intensity and fundamental frequency in our stimuli;
329 nevertheless, participants' judgments in the duration discrimination task were significantly
330 influenced by the stimulus frequency. Second, the quantitative comparison of duration and
331 intensity judgments demonstrates that the fundamental frequency contribution to intensity
332 judgments is not sufficient to fully account for its influence on perceived sound duration.
333 Finally, as shown by the rather low deviance reduction by the models, the discrimination
334 tasks were quite difficult for the participants. Still, they were able—in a statistical sense—to

335 perform the discriminations, and their judgments were robustly influenced by most signal

336 properties in expected directions (with the exception of f_o dynamics).

337 Stimulus design parameters necessarily have an effect on the response patterns: larger

338 variation in a primary acoustic correlate of a perceptual dimension (e.g. intensity level in

339 loudness discrimination) makes the task easier. The differences between the two stimuli in

340 a trial (determined by standard deviations of parameter distributions) must be detectable

341 but not so large as to make the stimuli qualitatively incompatible with each other. Larger

342 standard deviations in the complex stimuli would make a less difficult task and result in less

343 guessing from participants, reducing the precision of the model estimates; however, a task

344 that is entirely guessing is meaningless. Extreme values of the stimulus parameters could

345 result in different behavior than those for which participants were less sure, but this was not

346 the case in the current work (models using a subset of data within one standard deviation of

347 the standard parameter value showed the same significant effects and effect sizes).

348 Fitting the models for individual participants shows remarkable robustness and consistency

349 of the effects of signal attributes on both intensity and duration judgments. The effects of

350 signal duration and f_o level were significantly positive for *all* participants in both tasks, and

351 so were, naturally, the intensity effects in the intensity discrimination task. The effect of

352 signal intensity on duration discrimination judgment was significantly positive on the group

353 level (mixed effect model); on individual level it was primarily manifested by consistency of

354 effect direction (estimate sign) among the participants. In fact, this difference in statistical

355 patterns between the intensity and duration discrimination tasks, and, in particular, between

356 the f_o-level and intensity effects for duration discrimination provides further corroborative

357 evidence to our fundamental claim that the stimulus frequency effect on duration perception

358 is not solely mediated through the effect of fundamental frequency on perceived intensity of

359 the sound.

360 A somewhat more complex situation arises with participants' sensitivity to frequency dynamics.

361 The absolute dynamicity (the fact that one stimulus has greater range of f_o contour than the

362 other, regardless of the direction) had, at least for our stimuli, no significant and consistent

363 effect on participants' judgments in either task. The f_o slope direction also had no effect on

364 intensity judgments. For durational judgments, however, the (more) rising stimuli were judged

365 as longer than the (more) falling ones. Again, this effect was significant at the group level;

366 presumably, this significance arose primarily through consistency among the participants in

367 the direction of influence (individually, the effect was mostly non-significant). This result,

368 including the relative lack of robustness, is generally consistent with the findings discussed in

369 the Introduction: the evidence regarding this phenomenon is not unanimous and the results

370 reported in literature suggest a degree of influence of participants' language background on

371 the effect of stimulus dynamicity on durational judgments.

372 It is remarkable how well the participants were able to selectively attend to duration or

373 intensity in making their task-specific judgments. This requires an ability to decompose

374 the signal and can be quantified from the present measurements by comparing the relative

375 influence of acoustic dimensions across tasks. Indeed, stimulus duration had 4.5 times

376 stronger impact on duration judgments than loudness judgments (computed as the ratio of

377 estimates in the statistical models). Similarly, the stimulus intensity had 4.7 times stronger

378 impact on loudness judgments than on duration judgments. Hence, there is a symmetry in

379 the relative impact of intensity and duration on perceived loudness and length.

380 The findings presented here must be taken into account by the energy integration models

381 of the peripheral auditory system. The influence of fundamental frequency on perceived

382 duration, and more importantly, the lack of interaction between frequency and intensity

383 suggest an independent encoding of these signal characteristics, at least when used for

384 durational judgments. In general, the observed pattern fits well with the dual klepsydra

385 model for duration discrimination (34), assuming that the number of neural spikes entering

386 the dual klepsydra system depends, in possibly interconnected but cumulative way, on both

387 the intensity and fundamental frequency of the sound. By design, the spectral energy was

388 concentrated on a narrow area well separated from the fundamental frequency. The individual

389 harmonics are then indiscernible and allow for the frequency band of the signal energy and

390 the frequency of the pitch to be almost independent of each other.

391 In the current experiment, artificial stimuli were used with monolingual Finnish speakers, but

392 the effect size coefficients indicate that these participants show effects in the same direction:

393 increasing the intensity or fundamental frequency of a sound results in a perceived lengthening.

394 Similar experimental data (of the duration discrimination task only) from speakers of other

395 languages reported by (30) show qualitatively identical patterns for participants with different

396 language backgrounds. Interestingly, the relative sensitivity to individual signal characteristics

397 significantly differs for participants with different native languages, suggesting a degree

398 of plasticity in auditory processing. Despite these quantitative differences, both results

399 support the existence of general auditory biases and justify a generalization of the pattern of

400 fundamental frequency influencing duration judgment independently of an intensity confound.

401 Previous research has questioned whether the subtle effects of these biases have any importance

402 for human communication, namely, whether the well documented correlations between

403 frequency patterns and duration of syllables, associated in many languages with phonological

404 phenomena such as tone or quantity, arise primarily from production or perception constraints

405 (28, 44). The findings reported here yield some support to the importance of taking the

406 properties of auditory processing, even the subtle ones, into account when investigating

407 phonetic characteristics of different languages and their origins. Moreover, the precise nature

408 of these interplays partly determines the relevant, perceptually grounded degrees-of-freedom

409 space in signal characteristics available for linguistic communication.

410 **V. CONCLUSIONS**

411 Intensity, fundamental frequency, and duration of a sound influence its perceived intensity

412 (loudness) and duration. The contributions of these signal characteristics to perceived duration

413 and loudness are mutually independent and differ quantitatively between the two tasks of

414 intensity and duration discrimination. At least for the material used in this study, fundamental

415 frequency dynamics do not contribute consistently to either durational or loudness judgments.

416 Fundamental frequency contribution to duration perception cannot be fully attributed to

417 its effect on signal intensity, i.e, stimulus intensity alone cannot be responsible for duration

418 judgment effects.

419 **ACKNOWLEDGMENTS**

420 Gratitude to Nora Fontell for helping in data collection.

APPENDIX: DISCRIMINATION BY INDIVIDUAL SUBJECTS

The estimates and significances of duration difference, intensity difference, f_o difference and dynamic range difference (in two versions) effects obtained by fitting (simple) logistic regression models for the data for each individual speaker separately. The results for intensity discrimination task are shown in Appendix Tables 1 and 2, those for the duration discrimination tasks in Appendix Tables 3 and 4. The models reported in Appendix Tables 1 and 3 use the difference in absolute values of dynamic f_o range as a independent variable, those shown in Appendix Tables 2 and 4 use instead the alternative variable computed as a difference in raw range values.

Table 1: Effect estimates and significances of the logistic models effects model fitted for the individual subjects for *intensity* discrimination (difference between absolute frequency ranges). Significance codes: $0 <$***$< 0.001 <$**$< 0.01 <$*< 0.05.

Subject	interc.	dur. diff.	int. diff.	freq. diff.	ran. diff. (abs)
1	0.172	6.433 ***	0.199 ***	-0.025	-0.021
2	0.591 ***	4.866 **	0.201 ***	0.189 ***	0.067
3	-0.307 *	4.907 **	0.481 ***	0.068 *	0.033
4	0.171	4.375 ***	0.153 ***	0.068 *	0.026
5	0.232	6.261 **	0.538 ***	0.284 ***	0.192 ***
6	0.319 *	8.787 ***	0.355 ***	0.233 ***	-0.057
7	0.352 *	12.199 ***	0.270 ***	0.401 ***	-0.022
8	-0.187	4.372 *	0.661 ***	0.011	0.046
9	-0.113	5.126 ***	0.163 ***	0.053 *	-0.040
10	0.180	7.067 ***	0.204 ***	0.157 ***	0.052
11	0.174	8.685 ***	0.647 ***	0.128 ***	0.036

Table 2: Effect estimates and significances of the logistic models effects model fitted for the individual subjects for *intensity* discrimination (difference between absolute frequency ranges). Significance codes: $0 <^{***}< 0.001 <^{**}< 0.01 <^{*}< 0.05$.

Subject	interc.	dur. diff.	int. diff.	freq. diff.	ran. diff. (raw)
1	0.180	6.487 ***	0.201 ***	-0.030	-0.034
2	0.594 ***	4.781 **	0.204 ***	0.190 ***	-0.064 *
3	-0.308 ***	5.006 **	0.480 ***	0.067 *	0.001
4	0.183	4.161 **	0.161 ***	0.070 **	0.064 *
5	0.258	6.245 **	0.505 ***	0.283 ***	0.006
6	0.300	9.354 ***	0.353 ***	0.240 ***	-0.038
7	0.341 *	12.196 ***	0.271 ***	0.401 ***	0.016
8	-0.203	4.336 *	0.662 ***	0.002	-0.026
9	-0.115	5.176 ***	0.166 ***	0.054 *	0.033
10	0.185	7.145 ***	0.209 ***	0.151 ***	-0.030
11	0.185	8.705 ***	0.642 ***	0.129 ***	0.003

Table 3: Effect estimates and significances of the logistic models effects model fitted for the individual subjects for *duration* discrimination (difference between absolute frequency ranges). Significance codes: $0 <^{***}< 0.001 <^{**}< 0.01 <^{*}< 0.05$.

Subject	interc.	dur. diff.	int. diff.	freq. diff.	ran. diff. (abs)
1	2.141 ***	34.015 ***	0.043	0.338 ***	0.216 **
2	0.786 ***	24.618 ***	0.062	0.141 ***	0.002
3	-0.225	42.528 ***	0.092	0.111 **	0.043
4	-0.117	18.295 ***	0.044	0.084 **	0.067
5	-0.140	47.741 ***	0.123 *	0.390 ***	-0.045
6	0.525 **	27.480 ***	0.073	0.169 ***	0.046
7	1.147 ***	20.356 ***	0.100 **	0.368 ***	0.103
8	0.173	29.494 ***	0.068	0.150 ***	-0.003
9	0.861 ***	18.146 ***	0.008	0.138 ***	-0.024
10	0.618 ***	23.557 ***	0.035	0.169 ***	-0.046
11	-0.327	41.548 ***	0.189 ***	0.144 ***	-0.024

Table 4: Effect estimates and significances of the logistic models effects model fitted for the individual subjects for *duration* discrimination (difference between absolute frequency ranges). Significance codes: $0 <^{***}< 0.001 <^{**}< 0.01 <^{*}< 0.05$.

Subject	interc.	dur. diff.	int. diff.	freq. diff.	ran. diff. (raw)
1	1.993 ***	32.261 ***	0.023	0.317 ***	-0.002
2	0.773 ***	24.821 ***	0.062	0.140 ***	-0.039
3	-0.275	44.307 ***	0.085	0.126 **	0.119 **
4	-0.094	18.270 ***	0.033	0.086 **	0.022
5	-0.170	47.587 ***	0.124 *	0.392 ***	0.016
6	0.541 **	27.589 ***	0.077	0.169 ***	0.017
7	1.164 ***	20.255 ***	0.091 *	0.367 ***	0.016
8	0.129	29.669 ***	0.071	0.147 ***	0.055
9	0.859 ***	18.090 ***	0.010	0.135 ***	0.020
10	0.593 ***	23.935 ***	0.034	0.175 ***	0.054
11	-0.328	43.165 ***	0.195 ***	0.156 ***	0.145 **

REFERENCES

[1] D. M. Eagleman and V. Pariyadath, "Is subjective duration a signature of coding efficiency?", Philos. Trans. R. Soc. Lond., B, Biol. Sci. **364**, 1841–1851 (2009).

[2] H. Eisler, "Experiments on subjective duration 1868-1975: A collection of power function exponents", Psychol. Bull. **83**, 1154 (1976).

[3] R.-K. Cheng, C. J. MacDonald, and W. H. Meck, "Differential effects of cocaine and ketamine on time estimation: implications for neurobiological models of interval timing", Pharmacol. Biochem. Behav. **85**, 114–122 (2006).

[4] J. I. Lake, K. S. LaBar, and W. H. Meck, "Hear it playing low and slow: How pitch level differentially influences time perception", Acta Psychol. **149**, 169–177 (2014).

[5] B. Berglund, U. Berglund, G. Ekman, and M. Frankehaeuser, "The influence of auditory stimulus intensity on apparent duration", Scand. J. Psychol. **10**, 21–26 (1969).

[6] H. Burghardt, *Zusammenhang zwischen subjektiver und objektiver Dauer von Schallen (The relationship between subjective and objective sound duration)* (Dissertation, Technical University of Munich) (1972).

[7] I. Lehiste, "Influence of fundamental frequency pattern on the perception of duration", J. Phon. **4**, 113–117 (1976).

[8] S. Rosen, "The effect of fundamental frequency patterns on perceived duration", Speech Transmission Laboratory—Quarterly Progress and Status Report 17–30 (1977).

[9] J. Y. Jeon and F. R. Fricke, "Duration of perceived and performed sounds", Psychol. Music **25**, 70–83 (1997).

[10] B. C. Moore, *An introduction to the psychology of hearing (6th Edition)* (Emerald Group) (2012).

[11] S. Goldstone, W. T. Lhamon, and J. Sechzer, "Light intensity and judged duration", Bull. Psychon. Soc. **12**, 83–84 (1978).

[12] P. J. Kraemer, R. W. Brown, and C. K. Randall, "Signal intensity and duration estimation in rats", Behav. Process. **34**, 265–268 (1995).

[13] T. H. Rammsayer and M. Verner, "The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction", J. Vis. **14**, 17 (2014).

[14] G. Ekman, M. Frankenhaeuser, B. Berglund, and M. Waszak, "Apparent duration as a function of intensity of vibrotactile stimulation", Percept. Motor Skill **28**, 151–156 (1969).

[15] J. L. Santee and D. L. Kohfeld, "Auditory reaction time as a function of stimulus intensity, frequency, and rise time", Bull. Psychon. Soc. **10**, 393–396 (1977).

[16] D. L. Kohfeld, J. L. Santee, and N. D. Wallace, "Loudness and reaction time: Ii identification of detection components at different intensities and frequencies", Percept. Psychophys. **29**, 550–562 (1981).

[17] P. J. Kraemer, C. K. Randall, and R. W. Brown, "The influence of stimulus attributes on duration matching-to-sample in pigeons", Anim. Learn. Behav. **25**, 148–157 (1997).

[18] M. Epstein and M. Florentine, "Reaction time to 1-and 4-khz tones as a function of sensation level in listeners with normal hearing", Ear Hearing **27**, 424–429 (2006).

[19] D. McFadden, "Duration- intensity reciprocity for equal loudness", J. Acoust. Soc. Am. **57**, 702–704 (1975).

[20] C. D. Bonn and J. F. Cantlon, "The origins and structure of quantitative concepts", Cognitive Neuropsych. **29**, 149–173 (2012).

[21] B. Chandrasekaran, E. Skoe, and N. Kraus, "An integrative model of subcortical auditory plasticity", Brain Topogr. 1–14 (2013).

476 [22] R. Cumming, "The effect of dynamic fundamental frequency on the perception of
477   duration", J. Phon. **39**, 375–387 (2011).

478 [23] J. Gandour, "On the interaction between tone and vowel length: Evidence from thai
479   dialects", Phonetica **34**, 54–65 (1977).

480 [24] Q.-M. Kong, "Influence of tones upon vowel duration in cantonese", Lang. Speech **30**,
481   387–399 (1987).

482 [25] K. Suomi, "Temporal conspiracies for a tonal end: Segmental durations and accentual f0
483   movement in a quantity language", J. Phon. **33**, 291–309 (2005).

484 [26] M. Vainio, J. Järvikivi, D. Aalto, and A. Suni, "Phonetic tone signals phonological
485   quantity and word structure", J. Acoust. Soc. Am. **128**, 1313–1321 (2010).

486 [27] P. Lippus, E. L. Asu, P. Teras, and T. Tuisk, "Quantity-related variation of duration,
487   pitch and vowel quality in spontaneous estonian", J. Phonetics **41**, 17–28 (2013).

488 [28] C. Gussenhoven and W. Zhou, "Revisiting pitch slope and height effects on perceived
489   duration", in *Interspeech 2013*, 1365–1369 (2013).

490 [29] H. Lehnert-LeHouillier, "A cross-linguistic investigation of cues to vowel length percep-
491   tion", J. Phon. **38**, 472–482 (2010).

492 [30] J. Šimko, D. Aalto, P. Lippus, M. Włodarczak, and M. Vainio, "Pitch, perceived duration
493   and auditory biases: Comparison among languages", in *The 18th International Congress*
494   *of Phonetic Sciences (ICPhS 2015)* (2015).

495 [31] H. Altmann, I. Berger, and B. Braun, "Asymmetries in the perception of non-native
496   consonantal and vocalic length contrasts", Second Lang. Res. **28**, 387–413 (2012).

497 [32] W. J. Matthews, "How does sequence structure affect the judgment of time? exploring a
498   weighted sum of segments model", Cogn. Psychol. **66**, 259–282 (2013).

499 [33] P. Joris, C. Schreiner, and A. Rees, "Neural processing of amplitude-modulated sounds",
500   Physiol. Rev. **84**, 541–577 (2004).

501 [34] M. Wittmann, "The inner sense of time: how the brain creates a representation of
502   duration", Nat. Rev. Neurosci. **14**, 217–223 (2013).

503 [35] C. J. Plack, A. J. Oxenham, and R. R. E. Fay, *Pitch: neural coding and perception*
504   (Springer Science & Business Media) (2006).

505 [36] W. L. Brigner, "Perceived duration as a function of pitch", Percept. Motor Skill **67**,
506   301–302 (1988).

507 [37] G. J. Dooley and B. C. Moore, "Duration discrimination of steady and gliding tones:
508   A new method for estimating sensitivity to rate of change", J. Acoust. Soc. Am. **84**,
509   1332–1337 (1988).

510 [38] C. Kaernbach, "Simple adaptive testing with the weighted up-down method", Percept.
511   Psychophys. **49**, 227–229 (1991).

512 [39] M. Cooke, *Modelling auditory processing and organisation* (Cambridge University Press)
513   (1993).

514 [40] B. C. Moore and A. Sek, "Discrimination of frequency glides with superimposed random
515   glides in level", J. Acoust. Soc. Am. **104**, 411–421 (1998).

516 [41] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models
517   using lme4", Journal of Statistical Software **67**, 1–48 (2015).

518 [42] J. Pinheiro and D. Bates, *Mixed-Effects Models in S and S-PLUS*, Statistics and
519   Computing (Springer New York) (2000), URL https://books.google.fi/books?id=
520   N3WeyHFbHLQC.

521 [43] J. Järvikivi, M. Vainio, and D. Aalto, "Real-time correlates of phonological quantity
522   reveal unity of tonal and non-tonal languages", PLoS ONE **5**, 1–10 (2010).

523 [44] Yu, A. C. 2010. Tonal effects on perceived vowel duration. *Laboratory Phonology* 10,

524   151–168.

525 [45] D. Pressnitzer and RD. Patterson, "Distortion products and the perceived pitch of

526   harmonic complex tones", Physiological and psychophysical bases of auditory function,

527   97–104 (2001).

528 [46] D. Aalto and J. Šimko and M. Vainio, "Language background affects the strength of the

529   pitch bias in a duration discrimination task", Proceedings of the Interspeech 2013: 14th

530   annual conference of the international speech communication association 14th annual

531   conference of the international speech communication association, 243–247, (Springer

532   New York)(2013).