Perceptual and Motor Skills, 2003, 96, 297-304.

ESTIMATION OF SPEAKERS' WEIGHT AND HEIGHT FROM SPEECH: A RE-ANALYSIS OF DATA FROM MULTIPLE STUDIES BY LASS AND COLLEAGUES[1,2]

JULIO GONZALEZ

University Jaume I of Castellon

Summary.—Research on identification of speakers' weight and height from their speech has yielded controversial results. Conclusions from a series of reports conducted by Lass and colleagues must be interpreted with caution because they are based on comparisons of only two types of data, mean of actual values and mean of estimated values. Cohen and others clearly demonstrated that this method of analysis overstated the accuracy of listeners' judgments. Following Van Dommelen's procedure, data from six of Lass's reports were re-analyzed with more appropriate statistics. Results of re-analysis show that (a) listeners are not very efficient guessing the weight or height of speakers, as only 14% of the estimations correlated significantly with actual values. (b) Judgments are highly consistent under different acoustic conditions, suggesting that listeners follow vocal stereotypes about the body size of speakers, even though these stereotypes are wrong.

Many researchers have examined the question of whether a listener can judge a speaker's physical characteristics from speech. Research mainly has focused on identification of speakers' sex (Schwartz & Rine, 1968; Lass, Hughes, Bowyer, Waters, & Bourne, 1976; Lass, Almerino, Jordan, & Wals, 1980; Bennett & Montero-Diaz, 1982), age (Hartman & Danauer, 1976; Linville & Fisher, 1985; Neiman & Applegate, 1990; Traunmüller & van Bezooijen, 1994), race (Lass, Mertz, & Kimmel, 1978; Lass, Almerino, Jordan, & Wals, 1980; Walton & Orlikoff, 1994), and body size, i.e., weight and height (see reviews by Bricker & Pruzansky, 1976; Kreiman, 1997).

Research on identification of a speaker's weight and height have been particularly controversial. An early series of studies performed by Lass and associates (Lass & Davis, 1976; Lass, Beverly, Niscosia, & Simpson, 1978; Lass, Barry, Reed, Walsh, & Amuso, 1979; Lass, DiCola, Beverly, Barbera, Henry, & Badali, 1979; Lass, Hendricks, & Iturriga, 1980; Lass, Kelley, Cunningham, & Sheridan, 1980; Lass, Phillips, & Bruchey, 1980; Lass, Scherbick, Davies, & Czarnecki, 1982) reported that listeners were capable of making accurate direct estimations of speakers' weights and heights from re-

[1]This work was supported by Fundació Caixa Castelló-Bancaixa and the University Jaume I of Castellon, Spain, Project PIA99-01.

[2]Address correspondence to Julio Gonzalez, Department of Basic and Clinical Psychology and Psychobiology, University Jaume I of Castellon, 12080-Castellon, Spain or e-mail (gonzalez@psb.uji.es).

corded speech samples under a variety of conditions. Subsequent studies from other laboratories reported negative results (Gunter & Manning, 1982; Kunzel, 1989). Recently Van Dommelen and Moxness (1995) presented significant correlations between actual and estimated heights and weights only for male speakers, while neither male or female listeners were able to estimate weight or height of female speakers.

The results of studies performed by Lass and colleagues must be interpreted with caution given the method of analysis employed. Each study's conclusions were based mainly on comparisons of only two values, the mean of actual values of speakers' weights and heights and the mean of weight and height estimations by a group of listeners. Lass and associates generally found relatively small differences between these mean values and concluded that listeners' judgments were accurate estimations of the speakers' actual weights and heights. However, Cohen, Crystal, House, and Neuburg (1980) found that this method of data reduction and analysis overstated the accuracy claimed for the estimations. The reason is that most of the discrepancy between true and estimated measurements will be absorbed by regression toward the mean, that is, the use of an "average difference" cancels out the deviations from the actual values: a negative difference on one estimation cancels a positive difference on another estimation. Cohen, *et al.* (1980) obtained a very small mean difference in an ingenious "experiment." A group of "listeners" was instructed to (1) think of any male speaker and (2) estimate the weight of that speaker. At the same time an experimenter collected the weights of 14 men selected at random. Thus, in this experiment no recordings of speech were used, but both mean values were close. The mean of the actual weights of the 14 males ("speakers") was 172.2 lb., and the mean of the imaginary estimations made by the group of "listeners" was 173.5 lb., a difference smaller than most of the results in the studies by Lass and colleagues. In studies of this type, measures such as the square root of the average squared difference or simply the mean absolute difference are more appropriate statistics. Van Dommelen (1993) re-analyzed the data from Lass, Phillips, and Bruchey (1980) and found that listeners were not able to identify accurately the weights and heights of speakers. Although the overall mean of estimations was close to the overall mean of actual values, a statistical re-evaluation of data starting from each individual speaker yielded nonsignificant correlation coefficients between actual and estimated heights and weights. Also, another finding was that listeners' judgments were highly consistent, even though they were wrong. Van Dommelen obtained positive and significant correlations between various combinations of estimated values under different conditions of filtered speech, concluding that listeners made their judgments according to vocal stereotypes regarding the speaker's body size. Re-evaluation of the data from Lass, Phillips, and Bruchey (1980) was

possible because although no tables of raw data were given, the authors included histograms that could be used by Van Dommelen to reconstruct the raw data. The heights of the bars in the histograms represented mean estimated heights and weights and actual heights or weights for each individual speaker. These bars were measured by Van Dommelen directly on the original page with an estimated accuracy of 0.2 mm (1 mm corresponded to roughly 2 in. for height and 4 lb. for weight). Given the scatter in the data, that accuracy was considered sufficient.

However, an important set of papers reported by Lass and associates remain unanalyzed by Van Dommelen's method (1993). Considering the conclusions reached in these papers were based mainly on the comparisons of overall means (estimated vs actual weights and heights), cautious interpretation is needed. As raw data could be reconstructed from the histograms provided in these reports, it would be useful to evaluate the hypothesis using Van Dommelen's procedure. At the same time, a re-evaluation starting from the individual speakers' data offers the opportunity of testing, using other sources, the finding of a high consistency of listeners' judgments on the speaker's body size.

In the present study data from the following six reports were re-analyzed: (1) Lass, Beverly, Niscosia, and Simpson (1978), which was an investigation of direct estimations of speakers' heights and weights assessing the importance of the sex of the speaker and listener; (2) Lass, DiCola, Beverly, Barbera, Henry, and Badali (1979), in which the effect of phonetic complexity on judgment of the speaker's height and weight was studied, and four kinds of stimuli were employed to test four levels of phonetic complexity: isolated vowels, monosyllabic word, bisyllabic word, and sentences; (3) Lass, Barry, Reed, Walsh, and Amuso (1979), in which the authors used three conditions: forward-played (normal) speech, backward-played speech, and time-compressed speech to study the importance of temporal features of speech in speaker height/weight estimation; (4) Lass, Kelley, Cunningham, and Sheridan (1980), in which speaker's height and weight from voiced and whispered speech were compared to evaluate the effect of vocal pitch; (5) Lass, Hendricks, and Iturriga (1980), in which the consistency of listeners' judgments of speakers' heights and weights were assessed by comparing the estimations made in four experimental sessions; (6) Lass, *et al.* (1982), in which the effect of speakers' attempts to disguise their voices on listeners' accuracy in judgments of height and weight was assessed, using five conditions: normal speaking (control condition), speech in which the speakers tried to sound much taller and much shorter, as well as much heavier and much lighter than they actually were. In the present study, only the control condition was re-analyzed to compare with actual heights and weights of speakers.

METHOD

Following the same procedure employed by Van Dommelen (1993), raw data were reconstructed from the histograms presented in the six reports. The heights of the histogram bars represented actual height or weight for each individual speaker and mean estimated heights or weights under each experimental condition studied in each paper. In all papers, histograms for male and female speakers were separately provided in the four histograms included in each report. These were weights of male speakers, weights of female speakers, heights of male speakers, and heights of female speakers. The bars were measured in mm directly on the original page magnified to improve the accuracy of the measures. The estimated accuracy was of 0.2–0.3 mm corresponding to roughly 0.2–0.3 in. for height and 0.3–0.4 lb. for weight. Reliability of bar measurements was calculated by means of a re-

TABLE 1

PEARSON CORRELATION COEFFICIENTS AND CONFIDENCE INTERVAALS AT 95% FOR ACTUAL AND ESTIMATED MEASURES AND BETWEEN PAIRS OF ESTIMATES OF SPEAKERS' WEIGHTS AND HEIGHTS UNDER DIFFERENT CONDITIONS IN SIX STUDIES CONDUCTED BY LASS

	Actual vs Estimated Measures			
	Weight		Height	
Speakers:	Males	Females	Males	Females
Lass, Beverly, Nicosia, & Simpson (1978)				
Male listeners (a)	.33 (<0,.72)	.24 (<0,.67)	.57 (.08,.84)	.17 (<0,.63)
Female listeners (b)	.11 (<0,.59)	.34 (<0,.73)	.67 (.24,.88)	.43 (<0,.72)
Lass, DiCola, Beverly, Barbera, Henry, & Badali (1979)				
Vowels (a)	.72 (.31,.90)	.53 (.00,.83)	.51 (<0,.82)	.05 (<0,.57)
Monosyllabic words (b)	.37 (<0,.75)	.57 (.06,.85)	.51 (<0,.82)	.32 (<0,.73)
Bisyllabic words (c)	.45 (<0,.79)	.65 (.18,.88)	.63 (.15,.87)	.33 (<0,.73)
Sentences (d)	.43 (<0,.78)	.74 (.34,.91)	.56 (.04,.84)	−.43 (−.78,0<)
Lass, Barry, Reed, Walsh, & Amuso (1979)				
Forward played (a)	.38 (<0,.75)	.10 (<0,.58)	.22 (<0,.66)	.06 (<0,.56)
Backward played (b)	−.32 (−.72, 0<)	−.03 (−.53,0<)	−.07 (−.56,0<)	.01 (<0,.52)
Time-compressed (c)	.24 (<0,.67)	− 27 (−.69,0<)	.02 (<0,.53)	.47 (<0,.79)
Lass, Kelley, Cunningham, & Sheridan (1980)				
Voiced (a)	.48 (<0,.81)	30 (<0,.72)	.50 (<0,.81)	.09 (<0,.59)
Whispered (b)	.71 (.29,.90)	.38 (<0,.76)	.19 (<0,.65)	.53 (.00,.83)
Lass, Hendricks, & Iturriga (1980)				
1st estimation (a)	.13 (<0,.60)	.14 (<0,.61)	.05 (<0,.55)	.48 (<0,.80)
2nd estimation (b)	.38 (<0,.75)	.10 (<0,.58)	.16 (<0,.62)	.43 (<0,.77)
3rd estimation (c)	.40 (<0,.76)	.08 (<0,.57)	.37 (<0,.74)	.36 (<0,.74)
4th estimation (d)	.32 (<0,.72)	.11 (<0,.59)	.07 (<0,.56)	−.04 (−.54,0<)
Lass, Scherbick, Davies, & Czarnecki (1982)				
Normal speech	.38 (<0,.77)	.30 (<0,.73)	.45 (<0,.80)	−.04 (−.580<)

measurement of histograms in one paper (Lass, Beverly, Niscosia, & Simpson, 1978); all test-retest correlations were better than .99.

Results

Pearson product-moment correlations between actual and estimated weights and heights were calculated across speakers of the same sex. Table 1 shows coefficients and 95% confidence intervals (CI) by speaker's sex for the different conditions included in the six studies re-analyzed. Confidence intervals were computed according to each sample size (pairs of scores) and using Fisher's z. From the 64 coefficients, only 9 (14%) were statistically significant, and 0 were not included in the confidence intervals: 6 for male and 3 for female speakers. In Lass, Beverly, Niscosia, and Simpson (1978) only for the heights of men was there found a significant correlation between ac-

TABLE 1 (Cont'd)

Pearson Correlation Coefficients and Confidence Intervaals at 95% For Actual and Estimated Measures and Between Pairs of Estimates of Speakers' Weights and Heights Under Different Conditions in Six Studies Conducted by Lass

	Pairs of Estimates			
	Weight		Height	
	Males	Females	Males	Females
(a) vs (b)	.87 (.65,.96)	.90 (.72,.97)	.88 (.67,.96)	.67 (.24,.88)
(a) vs (b)	.41 (<0,.68)	.90 (.71,.97)	.82 (.51,.94)	.55 (.03,.84)
(a) vs (c)	.50 (<0,.81)	.82 (.51,.94)	.76 (.38,.92)	.76 (.38,.92)
(a) vs (d)	.33 (<0,.73)	.87 (.63,.96)	.53 (.00,.83)	.13 (<0,.62)
(b) vs (c)	.89 (.68,.96)	.89 (.68,.96)	.90 (.71,.97)	.57 (.06,.85)
(b) vs (d)	.93 (.79,.98)	.94 (.82,.98)	.75 (.36,.92)	−.10 (−.60,0<)
(c) vs (d)	.87 (.63,.96)	.90 (.71,.97)	.78 (.43,.93)	.25 (<0,.69)
(a) vs (b)	−.11 (−.59,0<)	.08 (<0,.57)	.01 (<0,.52)	.−.27 (−.69,0<)
(a) vs (c)	.55 (.05,.83)	.65 (.21,.87)	.39 (<0,.75)	.53 (.02 82)
(b) vs (c)	.06 (<0,.56)	.27 (<0,.69)	.19 (<0,.64)	−.08 (−.57,0<)
(a) vs (b)	.83 (.54,.95)	.53 (.00,.83)	.67 (.22,.89)	.26 (<0,.69)
(a) vs (b)	.75 (.39,.91)	.95 (.85,.98)	.76 (.41,.92)	.90 (.72,.97)
(a) vs (c)	.81 (.51,.93)	.89 (.69,.96)	.48 (<0,.80)	.92 (.77,.97)
(a) vs (d)	.81 (.51,.93)	.95 (.85,.98)	.63 (.17,.86)	.59 (.11,.85)
(b) vs (c)	.93 (.80,.98)	.96 (.88,.99)	.56 (.07,.83)	.92 (.77,.97)
(b) vs (d)	.93 (.80,.98)	.97 (.91,.99)	.72 (.33,.90)	.71 (.31, 90)
(c) vs (d)	.95 (.85,.98)	.95 (.85,.98)	.28 (<0,.69)	.67 (.24, 88)

tual and estimated values by both male and female listeners. Computing discrepancies or absolute differences of actual minus estimated values, there is no significant difference between male and female listeners in guessing the weights of male speakers ($t_{14} = -0.74$, ns). These discrepancies (16.2 lb. for male listeners and 17.3 lb. for female listeners) are much larger than data reported by Lass, Beverly, Niscosia, and Simpson (1978), wherein nonabsolute values were averaged (3.19 lb. for male listeners and 3.44 lb. for female listeners). Also, there are no significant differences between male and female listeners in guessing the weights of female speakers ($t_{14} = -0.68$, ns), heights of male speakers ($t_{14} = 0.07$, ns), or heights of female speakers ($t_{14} = 0.37$, ns).

In Lass, DiCola, Beverly, Barbera, Henry, and Badali (1979) correlations between actual and estimated values under four kinds of phonetic complexity were significant for guessing the weights of males from vowels, the weights of females from monosyllabic and bisyllabic words and sentences, and the heights of males from bisyllabic words and sentences. According to the mean differences, the authors did not find an increase in listeners' accuracy from the simplest to the most complex stimuli. Also in our re-analysis, no increase was found. Discrepancies in guessing males' weights were not significant among the four kinds of phonetic complexity ($F_{3,39} = 1.16$, ns). Absolute discrepancies also did not show significant differences among the four kinds of complexity estimating heights of males ($F_{3,39} = 0.04$, ns) or females ($F_{3,39} = 2.29$, ns). These only reached significance for female weights ($F_{3,39} = 2.91$, $p < .05$) given differences between vowels and bisyllabic words (12.5 vs 10.1 lb.; $t_{13} = 2.61$, $p < .05$).

In Lass, Barry, Reed, Walsh, and Amuso (1979) no correlation of actual and estimated values reached significance. However, except for females' heights, absolute discrepancies showed significant differences among conditions due to an anomalous behavior of *backward-played condition*, which reached discrepancies much greater—about twice those of the other conditions.

In Lass, Kelley, Cunningham, and Sheridan (1980) only estimations of males' weights from whispered speech reach a significant correlation with true values. Absolute discrepancies did not show significant difference across conditions (voiced vs whispered speech).

Finally, in Lass, Hendricks, and Iturriga (1980) and Lass, *et al.* (1982) no correlation between actual and estimated heights and weights reach significance regarding male or female speakers.

An interesting finding in the re-evaluation of the data from Lass, Phillips, and Bruchey (1980) made by Van Dommelen (1993) was the high consistency of estimations across different stimulus conditions (i.e., unfiltered, low-pass filtered, and high-filtered speech). In the present study this finding was tested by evaluating correlations between pairs of estimations from each

report (see Table 1). The overall picture is very different: of a total of 68 correlations 49 (72%) were significant. An important portion of the nonsignificant correlations corresponded to estimations made from backward-played speech and by Lass, Barry, Reed, Walsh, and Amuso (1979). These were different from those made in the other conditions, for those were forward-played and time-compressed speech.

Conclusions

In contrast to the conclusions of Lass and colleagues on the basis of overall means, the use of more appropriate statistics showed that listeners are not accurate in making their estimations of weights or heights of speakers when presented only their voices. Only 14% of the estimations resulted in significant correlations with the actual weights and heights of speakers. Also, when absolute differences between estimated and actual values are considered, the discrepancies are much larger than those reported by Lass and colleagues, wherein positive differences were cancelled by negative differences.

Re-analysis of the data from Lass's six reports supports the finding of Von Dommelen (1993), that listeners are very consistent in their judgments. This consistency occurs under different conditions of speech, whether normal voiced, whispered, filtered, time-compressed, or using vowels, words, or sentences as stimuli. The one exception is for backward-played speech, which is a case in itself. Listeners probably are guided, although incorrectly, by vocal stereotypes concerning the body size of speakers. In this sense, a clear example of an incorrect stereotype is listeners' use of the fundamental frequency of speech as a perceptual cue of the speaker's body size (Fitch, 1994). The data show instead that there is no significant correlation between the voice fundamental frequency and height or weight of human speakers (Lass & Brown, 1978; Kunzel, 1989). Another possible incorrect perceptual cue used to guess body size is formant frequencies of speech—or vocal tract size Fitch (1994) found that listeners associated synthesized vowels of lower formants with larger body sizes of the hypothetical speakers. However, the data currently available do not indicate a relation between formant frequencies and body size in adult humans of the same sex. In contrast to Fitch's findings (1997), in rhesus macaques, the few reports that have tangentially computed correlations within sex between formant frequencies—or derived measures as formant dispersion or vocal tract length—and human body size find nonsignificant or very weak coefficients (Van Dommelen & Moxness, 1995; Collins, 2000).

REFERENCES

Bennett, S., & Montero-Diaz, L. (1982) Children's perception of speaker sex. *Journal of Phonetics*, 10, 113-121.

Bricker, P. D., & Pruzansky, S. (1976) Speaker recognition. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press. Pp. 295-326.

Cohen, J. R., Crystal, T. H., House, A. S., & Neuburg, E. P. (1980) Weighty voices and shaky evidence: a critique. *Journal of the Acoustical Society of America*, 68, 1884-1886.

Collins, S. A. (2000) Men's voices and women's choices. *Animal Behaviour*, 60, 773-780.

Fitch, T. W. (1994) Vocal tract length perception and the evolution of language. Unpublished Ph.D. dissertation, Brown Univer.

Fitch, T. W. (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, 102, 1213-1222.

Gunter, C. D., & Manning, W. H. (1982) Listener estimations of speaker height and weight in unfiltered and filtered conditions. *Journal of Phonetics*, 10, 251-257.

Hartman, D. E., & Danauer, J. L. (1976) Perceptual features of speech for males in four perceived age decades. *Journal of the Acoustical Society of America*, 69, 713-715.

Kreiman, J. (1997) Listening to voices: theory and practice in voice perception research. In K.

Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing*. San Diego, CA: Academic Press. Pp. 85-108.

KUNZEL, H. J. (1989) How well does average fundamental frequency correlate with speaker height and weight? *Phonetica*, 46, 117-125.

LASS, N. J., ALMERINO, C. A., JORDAN, L. F., & WALS, J. M. (1980) The effect of filtered speech on speaker race and sex identification. *Journal of Phonetics*, 8, 101-112.

LASS, N. J., BARRY, P. J., REED, R. A., WALSH, J. M., & AMUSO, T. A. (1979) The effect of temporal speech alterations on speaker height and weight identification. *Language and Speech*, 22, 163-171.

LASS, N. J., BEVERLY, A. S., NISCOSIA, D. K., & SIMPSON, L. A. (1978) An investigation of speaker height and weight identification by means of direct estimations. *Journal of Phonetics*, 6, 69-76.

LASS, N. J., & BROWN, W. S. (1978) Correlational study of speakers' heights, weights, body surface areas, and speaking fundamental frequencies. *Journal of the Acoustical Society of America*, 63, 1218-1220.

LASS, N. J., & DAVIS, M. (1976) An investigation on speaker height and weight identification. *Journal of the Acoustical Society of America*, 60, 700-703.

LASS, N. J., DICOLA, G. A., BEVERLY, A. S., BARBERA, C., HENRY, G. H., & BADALI, M. K. (1979) The effect of phonetic complexity on speaker height and weight identification. *Language and Speech*, 22, 297-309.

LASS, N. J., HENDRICKS, C. A., & ITURRIGA, M. A. (1980) The consistency of listeners' judgments in speaker height and weight identification. *Journal of Phonetics*, 8, 439-448.

LASS, N. J., HUGHES, K. R., BOWYER, M. D., WATERS, L. T., & BOURNE, V. T. (1976) Speaker sex identification from voiced, whispered and filtered isolated vowels. *Journal of the Acoustical Society of America*, 59, 675-678.

LASS, N. J., KELLEY, D. T., CUNNINGHAM, C. M., & SHERIDAN, K. J. (1980) A comparative study of speaker height and weight identification from voiced and whispered speech. *Journal of Phonetics*, 8, 195-204.

LASS, N. J., MERTZ, P. J., & KIMMEL, K. L. (1978) The effect of temporal speech alterations on speaker race and sex identification. *Language and Speech*, 21, 279-290.

LASS, N. J., PHILLIPS, J. K., & BRUCHEY, C. A. (1980) The effect of filtered speech on speaker height and weight identification. *Journal of Phonetics*, 8, 91-100.

LASS, N. J., SCHERBICK, K. A., DAVIES, S. L., & CZARNECKI, T. D. (1982) Effect of vocal disguise on estimations of speakers' heights and weights. *Perceptual and Motor Skills*, 54, 643-649.

LINVILLE, S. E., & FISHER, H. B. (1985) Acoustic characteristics of perceived versus actual vocal age in controlled phonation by adult females. *Journal of the Acoustical Society of America*, 78, 40-48.

NEIMAN, G. S., & APPLEGATE, J. A. (1990) Accuracy of listener judgments of perceived age relative to chronological age in adults. *Folia Phoniatrica*, 42, 327-330.

SCHWARTZ, M. F., & RINE, H. E. (1968) Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America*, 44, 1736-1737.

TRAUNMÜLLER, H., & VAN BEZOOIJEN, R. (1994) The auditory perception of children's age and sex. In *Proceedings ICSLP94. 1994 International Conference on Spoken Language Processing, Yokohama, Japan*. Vol. 3. Pp. 1171-1174.

VAN DOMMELEN, W. A. (1993) Speaker height and weight identification: reevaluation of some old data. *Journal of Phonetics*, 21, 337-341.

VAN DOMMELEN, W. A., & MOXNESS, B. H. (1995) Acoustic parameters in speaker height and weight identification: sex-specific behaviour. *Language and Speech*, 38, 267-287.

WALTON, J. H., & ORLIKOFF, R. F. (1994) Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research*, 37, 738-745.

Accepted December 23, 2002.