Transworld Research Network
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India

Recent Res. Devel. Applied Phys., 9(2006): ISBN: 81-7895-213-0

Research in acoustics of human speech sounds: Correlates and perception of speaker body size

Julio González
Departament of Basic and Clinical Psychology, and Psychobiology University
Jaume I 12071- Castellón de la Plana, Spain

Abstract

If we consider people of different ages and sex, we can observe significant correlations between some acoustic measures of speech and the body size of the speaker. However, data become less clear when individual differences are studied and controlled by age and sex variables. This chapter reviews possible associations between some speech acoustic parameters and speaker body size within sex in human adults, avoiding sex- and age-related confounds. Also, it examines the ability of listeners to perceive speaker body size from speech samples.

Correspondence/Reprint request: Dr. Julio González, Departament of Basic and Clinical Psychology, and Psychobiology. University Jaume I. 12071- Castellón de la Plana, Spain. E-mail: gonzalez@psb.uji.es

Introduction

Human speech is a multidimensional signal that carries many kinds of information. In addition to the phonetic information, essential to language, the speech signal conveys a great deal of non-linguistic information about the speaker, including his/her sex, age, race, individual identity, dialect background, socio-economic status, personality, emotional state and other aspects.

The voice signal bears information about the speakers' physical characteristics. When we listen to an unknown talker over the telephone or radio, we rapidly develop a distinct impression of whether the talker is a man or a woman, an adult or a child [1]. This chapter focuses on a particular physical characteristic: the speaker body size. Specifically, we address the question of whether human speech sounds convey some information about the speaker's height and weight and whether listeners are able to perceive this information (Figure 1).

Figure 1. Do human speech sounds convey some information about speaker body size?

2. Speech sounds: Acoustic correlates of speaker body size

2.1. General considerations: between vs. within sex and age

If we consider people of different ages and sex, we can observe significant correlations between some acoustic measures of speech, such as the average fundamental frequency (*Fo*) and formant frequencies, and the body size of the speaker. As humans grow, their vocal folds also grow as their length and mass increase [2], which results in a lowering of *Fo*. The formants, or resonant frequencies of the supralaryngeal vocal tract, are dependent on the size of the vocal tube. As humans grow, the vocal tract length (VTL) also increases [3] and, in accordance with the source-filter theory [4], the formant frequencies of the vowels decrease [5]. Moreover, on average women are shorter and have smaller vocal folds and vocal tracts than men [3]. Consequently, adult men have the longest VTLs and the lowest formants, children have the shortest VTLs and the highest formants, while women have intermediate VTLs and formants [3]. Thus, significant correlations emerge between some acoustic measures and speaker body size, i.e. height or weight values, when the data are analyzed across different ages and sex.

However, data become less clear when individual differences are studied and controlled by age and sex variables. If we consider people of different ages (e.g., 6 - 30 years), we could find some relationship between intelligence and body size, for example; but it is known that intelligence and stature are not associated in the adult population. In other words, intelligence and body size are independent when age is controlled. In this chapter, we are interested in possible associations between some speech acoustic parameters and speaker body size within sex in human adults, avoiding sex- and age-related confounds.

2.2. Speaking fundamental frequency

In the past, a commonly acoustic parameter cited as a cue to body size was mean fundamental frequency (Fo) of voice [6-8]. Indeed, Fo shows a reliable negative correlation with body size across sex and age groups: Fo is higher in children and females and lower in males (on average, with taller and heavier bodies). However, when age and sex variables are controlled, data has repeatedly shown that correlation within sex between Fo and body size in adult humans is null or very weak [9-14]. The largest sample of speakers was studied by Künzel [10] within the forensic field. This author from the Bundeskriminalamt (Germany) calculated correlations between average Fo values of 105 male and 78 female adults and their individual height and weight measurements. No significant coefficient between the acoustic and physical parameters was found for the male or female group. Recently, Rendall and colleagues [14] carried out a study on human vowels and vowel-like grunts from baboons, a species chosen for its phylogenetic proximity to humans, similar vocal production biology and voice acoustic patterns. For adults of both species, males were larger than females and had lower mean voice Fo. Further to this however, Fo variation did not trace body-size variation within sexes in humans.

Curiously, although *Fo* is a bad cue for body size within the same sex and age group, it is wrongly used by people as a perceptual indicator of body dimensions. For example, van Dommelen and Moxness [12] found no significant correlations between *Fo* of speech samples recorded from 15 men and 15 women and their actual heights and weights. However, they observed important correlations between *Fo* and the estimations of such body parameters made by a set of listeners. Concretely, low *Fo* values were (wrongly) taken by listeners to indicate a large speaker body size. Similar results have also been observed in other experiments [15-16].

2.3. Formant parameters

The physics of acoustic tubes tell us that short tubes result in high resonance frequencies, whereas large tubes result in low resonance frequencies. As seen above, the increase in age from child to adult is marked by a decrease in the resonances of the vocal tube, or formant frequencies [17-19, 5], and that there are differences in formant frequencies between males and females [20-21, 5]. In this sense, formants could serve as an acoustic correlate of body dimensions. Some authors, like Tecumseh Fitch and colleagues [22-24, 3], state that the use of formant frequencies as a cue to the body size of a vocalizer has played an important role in the evolution of language. According to Fitch, if a mechanism for estimating body size from formants existed in the time of our prelinguistic ancestors, this could have provided a preadaptative basis for vocal tract normalization in humans.

For Fitch, the most important parameter as an acoustic cue to body size is the formant dispersion, defined as the average distance between each adjacent pair of formants. In a study of vocalizations in rhesus macaques, Fitch [23] found that formant dispersion was closely linked to both VTL and body size. Fitch’s data revealed a link between VTL and body size as a result of a tight anatomical correlation between both elements. The author concluded that unlike Fo, formant dispersion could provide a robust cue to body size in most mammals.

In contrast to Fitch’s findings in rhesus macaques however, currently available studies on human beings have found only a weak link between formant parameters and body size within adults of the same sex [12, 14, 15, 25]. Van Dommelen and Moxness [12] investigated the ability to judge speaker body size from Norwegian speech samples. In addition to body measurements, several acoustic correlates were considered as independent variables. Regression analysis involving Fo, formant frequencies, spectrum energy below 1 kHz, and speech rate yielded no significant correlations between these parameters and the speaker’s height and weight (the only exception being between the male speaker’s weight and speech rate). Collins [25] studied the relationship between male voice characteristics and female judgments about the speaker’s attractiveness. Body measurements (weight, height, and hip and shoulder width) and acoustic measures of five Dutch vowels uttered by 34 men (formant frequencies, overall peak and the first five harmonic frequencies) were included as independent variables. Data indicated that male voices with low-frequency harmonics were rated as being more attractive, but there was no relationship between any vocal and body characteristic. Moreover, regression coefficients between body measures and formant dispersion provided no significant results.

González [13] performed two experiments to investigate the relationship between formant parameters (frequencies and dispersions) and body size in human adults. In Experiment I, correlation coefficients were obtained between formant parameters from the five Spanish vowels uttered by 82 speakers and their heights and weights. In Experiment II, correlations were calculated from formants obtained by means of a long-term average analysis (LTAS) of connected speech uttered by 91 speakers. The analysis indicated that coefficients within sex from both experiments were null or quite weak. At the same time, the correlations within the female group were greater than in male group. In the study of Rendal *et al*. [14] on vowel-like baboon grunts and human vowels, the authors observed that the human formant variation correlated significantly with speaker height only in males (N=34) and not in females (N=34). Concretely, the strongest correlation was found for the fourth formant (F4) of schwa vowels. Owren & Anderson [26] used speech samples produced by 100 male athletes during televised interviews. They extracted Fo and formant measures from both running speech and filled pauses (i.e., “ah”, “um”) and these were placed in relation with individual statures. Multiple-regression analysis accounted for 17% of variance in speaker height at the most. Bruckert *et al*. [16] found that men with a low score in formant frequencies and dispersions tended to be taller but this result did not reach statistical significance.

In general, the correlations between formant parameters and body size in human adults of the same sex group are null or quite weak. Bearing in mind that formants are determined by the size and shape of the vocal tube, this suggests that the pattern of individual vocal tract development is relatively free from skeletal size constraints, maybe owing to the human descent of larynx from the standard mammal position. This disassociation of vocal-tract and body size is more important in human males.

2.4. Other parameters

Other parameters (speech rate, spectrum energy below 1 kHz, overall peak and the first five harmonic frequencies, etc.) have been considered as possible acoustic cues of speaker body size in some studies, but the results were not worthy of emphasis. Van Dommelen & Moxness incidentally obtained a significant correlation between male speaker weight and speech rate; i.e.

heavier males spoke at a slower rate. But this finding has not been replicated in other works [15-16].

Recently, González [27] calculated correlations between a set of 27 voice parameters [MDVP-Multi-Dimensional Voice Program, Kay Elemetrics Corp.] and 4 body measurements from 134 speakers of both sexes belonging to the same age group (20-29 years). All significant coefficients obtained within sexes were below 0.35.

3. Perception of body size from speech

We can view the perception of body size within a more general question of whether a listener can judge a speaker's physical characteristics from speech [28-29]. Research has mainly focused on the identification of speaker sex, age, race, and body size, i.e. weight and height. However, research on the identification of a speaker's weight and height has been particularly controversial.

3.1. Early research

An early series of studies performed by Norman J. Lass and his collaborators from the West Virginia University reported that listeners were capable of making accurate estimations of speakers' weights and heights from recorded speech samples under a variety of conditions.

Different tasks. In a first study published in 1976 [30], a standard prose passage was recorded by 30 speakers, 15 males and 15 females. The recorded readings were played to a group of 30 normal hearing persons who served as judges in two separate sessions. In one session, they were asked to determine the height of each speaker by means of a multiple-choice response sheet containing four choices: under 5 ft (152.4 cm); between 5 ft – 5 ft 5 in (152.4 – 165.1 cm); between 5 ft 6 in – 6 ft (167.6 – 182.9 cm); and over 6 ft. In another session, weight judgments were also made between four choices: under 100 lb (45.4 kg); between 100 lb – 150 lb (45.4 – 68.4 kg); between 151 lb – 200 lb (68.5 – 90.7 kg); and over 200 lb. Authors reported that the listeners were capable of identifying the heights of male and female speakers and the weights of male speakers with a slightly better than chance-guessing accuracy.

A second study [31] tested if listeners were capable of making accurate direct estimations of speaker's heights and weights from recorded speech samples. Instead of making four-choice responses, on this occasion the subjects were asked to make estimations in inches and pounds. Results showed that the average difference for all speakers and listeners between actual and estimated heights and weights was only 0.80 in and 3.48 lb, respectively. The authors concluded that participants were also capable of accurately identifying the approximate heights and weights of speakers in this more demanding task. Overall, listeners were more accurate in the identification of speaker heights than speaker weights. Moreover, the sex of the speakers and listeners did not significantly affect the identification judgments.

In a subsequent work, Lass and colleagues [32] found that listeners were able to make accurate comparative judgments of height and weight when presented with *pairs* of recorded speech samples. Subjects were asked to judge which speaker of each pair was taller or heavier. Discrimination responses resulted better than chance-guessing levels.

Lass *et al*., [33] studied the effect of *phonetic complexity* using four kinds of stimuli: isolated vowels, monosyllabic words, bisyllabic words and sentences. Results indicated that listeners accurately identified height and weight at all levels of phonetic complexity investigated. Furthermore, no regular progressive trend was evident in accuracy from the simplest (isolated vowels) to the most complex stimuli (sentences).

It seems that language *comprehension* does not play a major role in height and weight estimations. In an experiment, Lass and colleagues [34] created a total of three master tapes under three experimental conditions: a) American speakers' readings in English, b) foreign speakers' readings in their native languages, and c) foreign speakers' readings in English. These researchers found that the listeners' accuracy varied only slightly between an American reading in English and a foreign speaker reading in a foreign language, except for weight estimates of male speakers. Moreover, accuracy differences between the second and third conditions were statistically non-significant.

Stimulus Manipulation. One portion of research developed by Lass' group acoustically manipulated the stimuli in an attempt to isolate and define the acoustic cues in the voice which carried information about speaker body size. The purpose of an article published in 1979 [35] was to determine the importance of *temporal* features of speech in speaker height and weight identification. The readings of a standard passage by 30 speakers, 15 females and 15 males, were presented in three ways: forward-played (or normal), backward-played, and time-compressed. The last condition contained readings compressed to 40% of their original recording times. Data indicated that temporal alteration of the speech signal by means of backward-playing, where the normal sequence of articulatory events is disturbed, adversely affected listener judgments of height and weight. However, time compression of the speech signal appeared to have no significant effect on listener accuracy. While altering the temporal aspects of signal, time compression does not disturb the normal sequence of articulatory events.

In an attempt to determine the relative importance of portions of the broadband speech *spectrum* in speaker body size identification, Lass *et al*. [36]

used three kinds of stimuli: unfiltered (or normal), 255 Hz low-pass filtered voices, and 255 Hz high-pass filtered voices. Results indicated that accuracy was not significantly affected by filtering the speech signal.

In another work, Lass' group [37] studied the effect of vocal *pitch* on speaker height and weight identification by comparing results from voiced and whispered speech stimuli. Data demonstrated that identification accuracy did not differ significantly between both conditions, indicating that vocal pitch does not appear to play a major role in such identification tasks. All these results suggest that information on speaker body size carried by the speech signal is not located in a single acoustical parameter.

Finally, an experiment [38] was performed to determine the effect of speakers' attempts to disguise their voices to sound much taller or much shorter, as well as much heavier and much lighter, than they actually were. Results showed that, although a majority of speakers yielded height and weight estimates consistent with the intended disguise conditions, the differences among all conditions (four disguise and one normal condition) were relatively small.

3.2. Shortcomings of early research

Some subsequent studies from other laboratories did not support Lass' findings and reported negative results. Gunter & Manning [39] compared listener estimations of speaker heights and weights from unfiltered and filtered speech samples, and found that listeners were unable to make accurate estimations under both conditions. More recently, Van Dommelen and Moxness [12] obtained 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 for female speakers.

Unfortunately, the results of studies performed by Lass and associates must be interpreted with caution given the method of analysis employed. Conclusions were based mainly on comparisons of only two values, the mean of the actual values of speakers' weights/heights and the mean of weight/height estimations made by a group of listeners. Lass and colleagues generally found relatively small differences between these mean values and concluded that listeners' judgments were accurate estimations of such body parameters. In other words, the Lass approach was characterized by pooling data across groups of speakers and listeners rather than working with individual values.

Cohen and collaborators [40] demonstrated that this method of data reduction and analysis overstated the accuracy claimed for the estimations. The main reason is that most of discrepancy between true and estimated measurements is absorbed by regression toward the mean: a negative difference on one estimation cancels a positive difference on another. These authors demonstrated this effect by means of an ingenious “experiment”. A group of “listeners” was instructed to: a) think of any male speaker, and b) estimate the weight of that speaker. At the same time, an experimenter roamed through the laboratory and collected the weights of 14 men selected at random. No recordings of speech were used in this experiment, 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, an even smaller difference than most of the results obtained in the studies undertaken by Lass and colleagues

Van Dommelen [41] made a statistical reevaluation of the data presented in Lass *et al*. [36] from filtered and unfiltered voices. All correlations between individual values of actual weights and estimated weights under different filtering conditions yield statistically non-significant coefficients. The overall picture of speakers’ heights fully paralleled that of the estimations of weights. Van Dommelen concluded that the Lass *et al*’s [36] claim that listeners are able to accurately identify the heights and weights of speakers when presented with their voices only cannot be maintained.

Following this procedure, data from six of Lass’ reports were reevaluated by González [42] with appropriate statistics. Results of the reanalysis revealed that listeners were not so efficient at guessing the weight or height of speakers as only 14% of the estimations correlated significantly with actual values.

3.3. Recent research

The reanalysis of Lass’ data presented a discouraging picture on the listener capability to perceive speaker body size from speech samples. From the forensic field, Künzel [10] stated (p. 122): “At present it can rather be assumed that information on height and weight not only is not located in a single acoustical parameter —a fact acknowledged by Lass too— but that it is not contained in the speech signal at all.”

Despite this pessimistic statement, perception of speaker body size from speech is still an open question. In a recent work, González and Oliver [15] used a low-demanding task to demonstrate that listeners are capable of perceiving speaker height. Materials consisted of natural speech samples from male and female speakers with extreme heights, 20 short and 20 tall persons of each sex. Two classes of stimuli were used, a sentence and a sustained vowel pronounced by each speaker. A total of 75 listeners performed a tall/short binary decision task on the speaker of each speech sample. Results revealed that: 1) subjects were able to discriminate voices according to the speaker’s height, 2) sentences yielded a higher percentage of success (66.3%) than sustained vowels (54.3%, also significant), 3) no differences of judgment accuracy according to the speaker sex, although listeners (wrongly) felt that the task was easier with male voices, 4) the correlations between average Fo and formant frequencies with individual speaker heights were null or very weak in each sex group. The authors concluded that some information exists on speaker height that can be processed by the listeners, but this information is likely codified in the acoustical time-varying properties of speech and not located in a single acoustical feature. Figure 2 displays the results from the experiment carried out with the sentence stimuli; we can specifically see the distribution of listener responses (TALL vs. SHORT) in accordance with the speakers' sex and stature. It is evident that the success level was well above the chance level [$t(74)=10.16$, $p<.001$ for male speakers; $t(74)=10.40$, $p<.001$ for female speakers].

Rendal *et al.* [43] asked listeners to hear paired comparisons of the same short phrase spoken by two adults of the same sex, randomly paired in relation to height, and to indicate which was larger. Listeners of both sexes showed a modest but significant ability to identify the larger male (58.5%) but could not select the larger female (52.0%, non-significant).

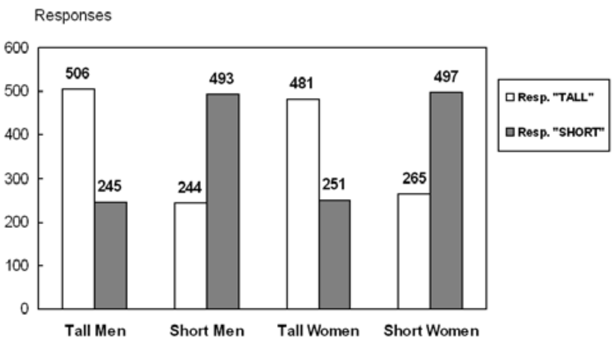

Figure 2. Distribution of responses "TALL" and "SHORT" given by subjects when listening to natural speech samples (a sentence) from Tall Men, Short Men, Tall Women, and Short Women. Adapted from González and Oliver [15].

Synthetic stimuli. In recent years, there is a trend to use computer-synthesized speech sounds to manipulate specific voice parameters in a controlled manner to observe the effect on the perceived body size.

One of the first studies was presented in the Doctoral Thesis of W.T. Fitch [22]. By means of a *Sensyn* speech synthesis package by Klatt [44], Fitch synthesized 8 vowel (schwa) sounds which had either a high or low fundamental frequency (100 or 150 Hz) and formant frequencies corresponding to one of four vocal tract lengths (15, 16, 17 or 18 cm). The vocal tract lengths chosen spanned the normal range for adult males. Each stimulus was presented six times to eleven subjects who had to circle a number between 1 and 7 to rate the apparent body size of the person producing the sound. Subjects found the task quite easy and straightforward. Nobody complained that it was difficult or impossible to do. Regression analysis revealed a significant correlation between formant frequencies and body size ratings. As expected, subjects associated low fundamental frequency (*Fo*) with larger body sizes and high *Fo* with smaller body sizes. The change in body size ratings between low and high *Fo* resulted in 2 units. The difference resulting from the formant manipulation was around 1 rating unit. However, there was no interaction between both acoustical variables: there was a clear effect of formant frequency which was completely independent of the effect of *Fo*.

Recently, a series of studies has been published by Roy D. Patterson's group from the University of Cambridge, using (re)synthesized speech sounds created by the STRAIGHT software [45-46]. In one experiment [47], a set of five English vowels were scaled to represent people with a wide range of fundamental frequencies, given by the GPR or glottal-pulse rate parameter, and vocal tract lengths (VTLs), including many which were well beyond the normal range of the population. Judgments of speaker size showed that, although they were affected by both VTL and GPR, the effect of VTL was stronger than that of GPR.

Listeners are also able to make fine discriminations between different perceived body sizes. Changes in simulated VTL in synthesized vowels of as little as 7% can be reliably discriminated [47]. Moreover, discrimination performance is much better when syllables are used rather than vowels alone. In this case, changes in simulated VTL of 4% can be discriminated [48].

Speech stimulus synthesis is based on the classical source-filter theory [4, 49], the reason why fundamental frequency is determined by the vibration rate of the simulated vocal folds (source), and formants are determined by the size and shape of the simulated vocal tract (filter). Thus, we can match the subject responses against the simulated VTL in experiments on body size perception. However, we cannot match responses against any real or simulated body size (height or weight). Obviously, we have no true body since the stimuli are artificial, but we cannot consider any simulated body size because a strong correlation between VTL and body size is not demonstrated. No study has proved a high VTL-body size correlation in human adults within a sex group. Any strong VTL-body size correlations found were across different age and sex groups [3]. An additional argument against such a high correlation is that correlations between formant frequencies (related with VTL) and body heights and weights in human adults are very weak within sex [13, 14, 50].

Consequently, in order to advance in the study of (true) speaker body size perception, we need to continue using natural stimuli in order to have real body sizes as a criterion to evaluate responses. Experiments with synthetic stimuli are valuable to know which perceptual cues are used by listeners to perceive body size, but they do not demonstrate that subjects make accurate estimations upon real body sizes. Indeed, data show that listeners follow vocal stereotypes regarding the body size of speakers, even though these stereotypes are wrong (e.g., the fundamental frequency).

3.4. Vocal stereotypes about speaker body size

One of Lass' papers [51] studied the consistency of the listeners judgments of speakers' heights and weights. The stimuli were presented to listeners in four experimental sessions to make direct estimations of such body parameters. Results indicated that listener estimations were consistent across all sessions.

When van Dommelen [41] made a reanalysis of the data presented by Lass *et al*. [36], although the estimations were not found to be altogether accurate, he observed that responses appeared to be highly consistent across different stimulus conditions. Van Dommelen concluded that speakers' voices contain features which are (incorrectly) used by listeners as indicators of body height and weight.

González [13] reanalyzed data from six of Lass' reports. Results indicated two details: 1) listeners had not been very efficient at guessing speaker body size, 2) however, judgments had been highly consistent under different speech conditions, whether normal voiced, whispered, filtered, time-compressed, or using vowels, words or sentences as stimuli. The one exception was for backward-played speech, which was a case in its own right.

All these and other data suggest that listeners are likely guided, although incorrectly, by vocal stereotypes regarding the body size of speakers. In this sense, a clear example of an incorrect stereotype is the listeners' use of the fundamental frequency of speech as a perceptual cue of the speaker's body size [12, 22]. Instead, the data show that there is no significant correlation between the voice fundamental frequency and the height or weight of adult speakers within each sex group [9, 10].

Another possible incorrect perceptual cue used to guess body size is formant frequencies of speech, or vocal tract size. As seen above, the data currently available do not indicate a strong relation between formant frequencies and body size in adult humans of the same sex.

4. Conclusions

In summary, according to the current data available, some conclusions can be stated:

1. Some acoustic features (*Fo* and formant parameters) of speech signal correlate with speaker body size across different ages and sex.

2. Contrary to common belief, the mean fundamental frequency (*Fo*) of voice shows a null or very weak relationship with body size of adult speakers of the same sex.
3. In contrast to Fitch's findings in macaque vocalizations, the relationship within sex between formant parameters and body size is quite weak in human adults.
4. Despite the null or weak correlations within sex groups, listeners use *Fo* and formants as perceptual cues of body dimensions. The consistency of their responses suggests that they are likely guided by vocal stereotypes regarding the body size of speakers.
5. Beyond *Fo* and mean formant parameters, speech signal conveys some information about speaker body size that listeners can exploit it. This information is not confined to a single parameter and is likely codified by means of some acoustical time-varying properties of speech, such as formant trajectories or others.

Acknowledgment

This work was partly supported by *Fundació Caixa Castelló-Bancaixa* and the Universitat Jaume I, Castellon (Project P1.1A2002-01) and the Ministry of Science and Technology of Spain (I+D+I, Project BSO2003-01002/PSCE).

References

1. Smith, R. R., and Patterson, R. D. 1975, J. Acoust. Soc. Am., 118, 3177.
2. Titze, I. R. 1975, J. Acoust. Soc. Am., 85, 1699.
3. Fitch, W. T., and Giedd, J. 1999, J. Acoust. Soc. Am., 106, 1511.
4. Fant, G. 1960, Acoustic theory of speech production, Mouton, The Hague.
5. Huber, J. E., Stathopoulus, E. T., Curione, G. M., Ash, T. A., and Johnson, K. 1999, J. Acoust. Soc. Am., 106, 1532.
6. Darwin, C. 1871, The descent of man and selection in relation to sex, Murray, London.
7. Morton, E. S. 1977, Am. Naturalist, 111, 855.
8. Laver, J., and Trudgill, P. 1979, Phonetic and linguistic markers in speech – Social markers in speech, S. Giles (Ed.), Cambridge Univ. Press, Cambridge, 1.
9. Lass, N. J., and Brown, W. S. 1978, J. Acoust. Soc. Am., 63, 1218.
10. Künzel, H. J. 1989, Phonetica, 46, 117.
11. Hollien, H., Green, R., and Massey, K. 1994, J. Acoust. Soc. Am., 96, 2646.
12. van Dommelen, W. A., and Moxness, B. H. 1995, Lang speech, 38, 267.
13. González, J. 2004, J. Phonetics, 32, 277.
14. Rendall, D., Kollias, S., Ney, C., and Lloyd, P. 2005, J. Acoust. Soc. Am., 117, 944.
15. González, J., and Oliver, J. C. 2004, Revista de Psicología General y Aplicada, 57, 21.

16. Bruckert, L., Lienard, J. S., Lacroix, A., Kreutzer, M., and Leboucher, G. 2006, Proc. Biol. Sci., 273, 83.
17. Eguchi, S., and Hirchs, I. J., 1969, Acta Otolaryngol. Suppl., 257, 1.
18. Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. 1995, J. Acoust. Soc. Am., 97, 3099.
19. Lee, S., Potamianos, A., and Narayanan, S. 1999, J. Acoust. Soc. Am., 105, 1455.
20. Peterson, G. E., and Barney, H. L. 1952, J. Acoust. Soc. Am., 24, 175.
21. Busby, P. A., and Plant, G. L. 1995, J. Acoust. Soc. Am., 97, 2603.
22. Fitch, T. W. 1994, unpublished Ph.D. thesis, Brown University.
23. Fitch, T. W. 1997, J. Acoust. Soc. Am., 102, 1213.
24. Fitch, T. W. 2000, Trends Cog. Sci., 4, 258.
25. Collins, S. A. 2000, Animal Behav., 60, 773.
26. Owren, M. J., and Anderson, J. D. 2005, J. Acoust. Soc. Am., 117, 2375.
27. González, J. 2006, manuscript submitted.
28. Bricker, P. D., and Pruzansky, K. 1976, Speaker recognition – Contemporary issues in experimental phonetics, N. J. Lass (Ed.), Academic Press, New York.
29. Kreiman, J. 1997, Listening to voices: theory and practice in voice perception research – Talker variability in speech processing, K. Johnson, and J. W. Mullennix (Eds.), Academic Press, San Diego.
30. Lass, N. K. and Davis, M. 1976, J. Acoust. Soc. Am., 60, 700.
31. Lass, N. J., Beverly, A. S., Nicosia, D. K., and Simpson, L. A. 1978, J. Phonetics, 6, 69.
32. Lass, N. J., Brong, G. W., Ciccolella, S. A., Walters, S. C., and Maxwell, L. 1980, J. Phonetics, 8, 205.
33. Lass, N. J., DiCola, G. A., Beverly, A. S., Barbera, C., Henry, G. H., and Badal, M. K. 1979, Lang. speech, 22, 297.
34. Lass, N. J., Riley, A. L., Giachetti, J. A., Cline, A. L., Dodson, C. K., and Bellotte, S. L. 1981, Percep. Mot. Skills, 53, 647.
35. Lass, N. J., Pamela, J. B., Reed, R. A., Walsh, J. M., Amuso, T. A. 1979, Lang speech, 22, 163.
36. Lass, N. J., Phillips, J. K., and Bruchey, C. A. 1980, J. Phonetics, 8, 91.
37. Lass, N. J., Kelley, D. T., Cunningham, C. M., and Sheridan, K. J. 1980, J. Phonetics, 8, 195.
38. Lass, N. J., Scherbick, K. A., Davies, S. L., and Czarnecki, T. D. 1982, Percept. Mot. Skills, 54, 643.
39. Gunter, C. D., and Manning, W. H. 1982, J. Phonetics, 10, 251.
40. Cohen, J. R., Crystal, T. H., House, A. S., and Neuburg, E. P. 1975, J. Acoust. Soc. Am., 68, 1884.
41. van Dommelen, W. A. 1993, J. Phonetics, 21, 337.
42. González, J. 2003, Percept. Mot. Skills, 96, 297.
43. Rendall, D., Vokey, J. R., Nemeth, C., and Ney, C. 2005, J. Acoust. Soc. Am., 117, 2372.
44. Klatt, D. H., and Klatt, L. C. 1990, J. Acoust. Soc. Am., 82, 820.
45. Kawahara, H., Masuda-Kasuse, L., and de Cheveigne, A. 1999, Speech Commun., 27, 187.
46. Kawahara, H., and Irino, T. 2004, Underlying principles of high-quality speech manipulation system STRAIGHT and its application to speech segregation –

Speech segregation by Humans and Machines, P. Divenyi, (Ed.), Kluver Academic, Boston.

47. Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H. and Irino, T. 1975, J. Acoust. Soc. Am., 117, 305.
48. Ives, D. T., Smith, D. R. R., and Patterson, R. D. 2005, J. Acoust. Soc. Am., 118, 3816.
49. Dudley, H. 1939, J. Acoust. Soc. Am., 11, 169.
50. Greisbach, R. 1999, Forensic Ling., 6, 1350.
51. Lass, N. J., Hendricks, C. A., and Iturriaga, M. A. 1980, J. Phonetics., 8, 439.