The frequency scale of speech intonation

Citation for published version (APA):
Hermes, D. J., & Gestel, van, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, *90*(1), 97-102. https://doi.org/10.1121/1.402397

DOI:
10.1121/1.402397

Document status and date:
Published: 01/01/1991

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 31. Mar. 2023

The frequency scale of speech intonation

Dik J. Hermes and Joost C. van Gestel
Institute for Perception Research/IPO, P.O. Box 513, NL 5600 MB Eindhoven, The Netherlands

(Received 8 May 1990; accepted for publication 3 January 1991)

In intonation research, prominence-lending pitch movements have either been described on a linear or on a logarithmic frequency scale. An experiment has been carried out to check whether pitch movements in speech intonation are perceived on one of these two scales or on a psychoacoustic scale representing the frequency selectivity of the auditory system. This last scale is intermediary between the other two scales. Subjects matched the excursion size of prominence-lending pitch movements in utterances resynthesized in different pitch registers. Their task was to adjust the excursion size in a comparison stimulus in such a way that it lent equal prominence to the corresponding syllable in a fixed test stimulus. The comparison stimulus and the test stimulus had pitches running parallel on either the logarithmic frequency scale, the psychoacoustic scale, or the linear frequency scale. In one-half of the experimental sessions, the test stimulus was presented in the low register, while the comparison stimulus was presented in the high register, and, conversely, for the other half of the sessions. The result is that, in all cases, stimuli are matched in such a way that the average excursion sizes in different registers are equal on the psychoacoustic scale.

PACS numbers: 43.71.Bp, 43.66.Hg, 43.71.Cq

INTRODUCTION

In physics, frequency is generally expressed in terms of the unit hertz (Hz). In various branches of hearing research, other units are used. In music perception, the relative distance between two tones is expressed in a musical interval such as the semitone and the octave. In this musical scale, equal distances represent equal frequency proportions, which amounts to using a logarithmic frequency scale. In psychoacoustics, the Mel scale has been used, based on a subjective measure of pitch "magnitude" (Stevens *et al.*, 1937; Stevens and Volkman, 1940). More often, a related frequency scale is used, the Bark scale. This is approximately linear for frequencies below 500 Hz and approximates a logarithmic frequency scale for higher frequencies. The Bark scale is derived from measurements of the frequency selectivity of the human auditory system, as measured by the so-called critical bandwidth (Fletcher, 1940; Zwicker *et al.*, 1957; Zwicker, 1961). Analytical expressions for the Bark scale are presented by Zwicker and Terhardt (1980), and by Traunmüller (1990). A newer variety of this scale is the equivalent-rectangular-bandwidth-rate (ERB-rate) scale (Patterson, 1976), for which analytical expressions are presented in Moore and Glasberg (1983) and Glasberg and Moore (1990). In the ERB-rate scale, the critical bands are narrower, especially at lower frequencies, than in the Bark scale. For frequencies below 500 Hz, the ERB-rate scale is neither linear, such as the Bark scale, nor logarithmic, but something in between. A detailed and quantitative discussion on this scale is given by Glasberg and Moore (1990).

The frequency scales derived from the frequency selectivity of the auditory system have been associated with distances along the basilar membrane (Fletcher, 1953; Greenwood, 1961, 1990). This has been verified anatomically for cat by Wilson and Evans (1977) and by Liberman (1982), who also supplied formulas relating frequency to place on the basilar membrane. In this study, the formulas from Greenwood (1961, 1990) for man were used:

$$E = 16.7 \log_{10}(1 + f/165.4), \tag{1}$$

$$f = 165.4(10^{0.06E} - 1), \tag{2}$$

where f is frequency in Hz, and E is the ERB-rate in ERB. These expressions give about the same values as the ERB-rate scale published by Moore and Glasberg (1983) (Moore and Glasberg, 1986, p. 254). Therefore, the psychoacoustic scale as used in this study will also be indicated with ERB-rate scale.

In intonation research, pitch of speech has either been expressed in Hz (e.g., Cooper and Sorensen, 1981) or in semitones (e.g., 't Hart *et al.*, 1990). In an experiment set up to find out whether prominence-lending pitch movements should be expressed in Hz or in semitones, Rietveld and Gussenhoven (1985) concluded that prominence judgments of their subjects were in better agreement with an Hz scale than with a scale of semitones. This result was based on comparisons of the prominence of syllables with pitch movements in different frequency regions. As their stimulus set comprised only sentences recorded from a female speaker, which were resynthesized in an equal or in a lower pitch register, they concluded that it would be premature to conclude that prominence-lending pitch movements should be expressed in Hz. According to Graddol (1986, p. 228), the pitch range used by most women seems to be rather less than that used by most men, when expressed in semitones, but larger, when expressed in a linear scale. Choosing between

0001-4966/91/070097-06$00.80

the two, he concludes that, "whenever intervals in pitch must be compared at different frequencies, a log scale is to be preferred." Only Traunmüller *et al.* (1989) have so far considered the possibility that pitch movements in speech intonation may best be expressed in a scale derived from the frequency selectivity of the auditory system. Following Graddol (1986), they showed preference for the logarithmic frequency scale, however.

This problem has consequences for various applications. If in synthetic speech, e.g., one wants to give the same prominence to accented syllables in male as in female speech, the excursion of the pitch movements must be the same on the frequency scale in which the prominence of pitch movements is perceived. As female and male voices differ by almost 1 oct, the difference between these approaches can cause considerable discrepancies. For example, an excursion of 120 to 180 Hz in a male voice would correspond to an excursion of 240 to 300 Hz in a female voice if a linear scale were used, whereas an excursion of 240 to 360 Hz would provide equal prominence if a logarithmic scale were used. On an ERB-rate scale, equal prominence would require an excursion of 240 to 325 Hz.

In order to decide on which scale the excursions of pitch movements are perceived, subjects adjusted the variable excursion size of a pitch movement in a comparison stimulus to the fixed excursion size of a pitch movement in a test stimulus resynthesized in a different frequency register. This was done both in sessions in which the test stimulus was in a low register, while the comparison stimulus was in a high register, and for sessions in which the test stimulus was in a high register, while the comparison stimulus was in a low register. Furthermore, this was done for six different excursion sizes, and for three different prominence-lending pitch movements, a rise, a rise–fall, and a fall.

I. EXPERIMENT

A. Materials

The stimuli consisted of modified versions of one utterance, /mamáma/, spoken by a male speaker. Its duration was 0.77 s. The second syllable carried an accent. Pitch modifications were applied with the pitch-synchronous overlap and add (PSOLA) technique (Hamon *et al.*, 1989), resulting in very natural sounding speech stimuli. Duration and amplitude relations were kept constant.

These stimuli were resynthesized with one of three prominence-lending pitch movements, a rise, a rise–fall, and a fall, superimposed on declination lines as displayed in Fig. 1. These pitch contours consisted of lines that were straight on either a linear frequency scale, an ERB-rate scale, or a logarithmic frequency scale, resulting in three different versions, which will be referred to as LIN, ERB, and LOG, respectively. All versions had a declination end point of either 75 or 180 Hz, defining the low versions and the high versions. The high versions sounded like a male falsetto voice.

All sessions consisted of adjustment runs in which two stimuli, a low and a high version, were repeatedly presented to the subject with an interstimulus interval of 1 s. The stimulus presented first will be referred to as the test stimulus. This was fixed within one run. The second stimulus, referred to as the comparison stimulus, was presented in another register and had a variable excursion size. In the first trial of a run, the excursion size in the comparison stimulus was zero. Subjects were asked first to increase the excursion size in the comparison stimulus to such an extent that the prominence of its accented syllable clearly exceeded the prominence of the accented syllable in the test stimulus. In the next trials, the subjects were asked to decrease and increase the excursion size in the comparison stimulus until it was judged to give the same prominence as the pitch excursion in the test stimulus. When the subject had done this, the next run started. In each session, there were six runs for six different excursion sizes in the test stimulus. The six different excursion sizes in the test stimulus were presented in random order with a different order in each session.

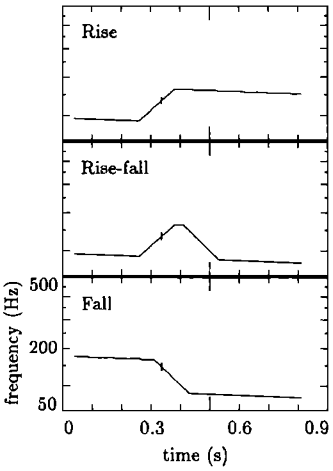

FIG. 1. The three different prominence-lending pitch movements. The pitch movements superimposed on the declination line give prominence to the second syllable, the vowel onset of which is indicated by the cross bar. The rise starts from low declination 70 ms before vowel onset and ends at high declination 50 ms after the vowel onset of the second syllable. The rising part of the rise–fall has the same timing and is followed by the falling part, which starts 80 ms and ends 200 ms after the vowel onset. The fall starts from high declination 20 ms before the vowel onset of the second syllable, and ends at low declination 100 ms after the vowel onset. The same timing is used in the IPO text-to-speech system.

As mentioned, the two stimuli matched in prominence were presented in different registers. In one set of sessions, the test stimuli were presented in the high register, while the comparison stimuli were presented in the low register (see Fig. 2). These will be referred to as downward sessions. In another set of sessions, the test stimuli were low in register, and the comparison stimuli high. These will be referred to as upward sessions (see Fig. 3). These upward and downward sessions took place for all three pitch movements and for each of the three different frequency scales, giving eighteen different sessions.

In each session, the comparison stimuli formed a set of ten stimuli with increasing excursion size. They were constructed in such a way that, within one register, they were almost identical in all three frequency scales [compare Fig. 2(b), (d), and (f), and Fig. 3(b), (d) and (f)]. Exact equality was impossible as the lines that made up the pitch con-

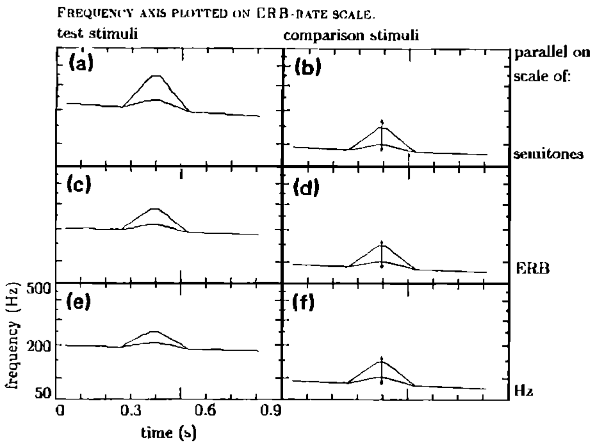

FIG. 2. Stimulus configuration of the downward adjustment sessions for the rise–fall. The range of the test stimuli is displayed in (a), (c), and (e), showing the test stimulus with the lowest and the highest excursion size. The range of the comparison stimuli is shown by the arrow in (b), (d), and (f). The continuous lines show the prominence-lending pitch movements, which, in the frequency scale mentioned at the right, run parallel to the displayed test stimuli. Notice that the comparison stimuli in the three frequency scales are very much the same, whereas the test stimuli are different in both the start frequency of the declination and in the size of the excursions. In anticipation of the result, all stimuli are presented with a ERB-rate scale as ordinate.

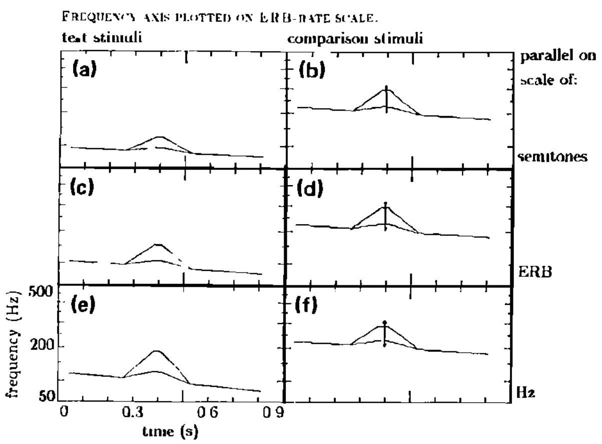

FIG. 3. Stimulus configuration of the upward adjustment sessions for the rise–fall; otherwise, as Fig. 2.

tour were required to be straight in one of the three frequency scales. A transformation of one scale to another being nonlinear, the pitch contours could only be straight in one frequency scale. The differences were very small, however. The low versions of these comparison stimuli had a common end point of their low declination line of 75 Hz and a start point of 93 Hz. This amounts to a declination of 4.85 semitones per second (st/s). This is derived from the rule for a declination of $11/(t+1.5)$ st/s, t being the duration of the utterance in this case 0.77 s. (This rule for the declination line is used in the Institute-for-Perception-Research diphone-speech-synthesis system. It dates from times when intonation was described in semitones at IPO.) For the high versions, the common end point was 180 Hz, and the start point was 223.3 Hz. The pitch movements were superimposed on these declination lines. The smallest excursion size was zero, and, for the low versions, the highest was 12.17 semitones for the LOG versions, 2.00 ERB for the ERB versions, and 76.5 Hz for the LIN versions. For the high versions, the largest excursion was 8.25 semitones for the LOG versions, 2.00 ERB for the ERB versions, and 109.9 Hz for the LIN versions. The intermediate excursion sizes were such that they were equidistant in the corresponding frequency scale. As mentioned, the result was such that the comparison stimuli in all three frequency scales were very much the same. The six test stimuli were then produced by an upward (for the downward sessions) or downward (for the upward sessions) pitch shift of the six middle versions of these comparison stimuli in such a way that the end points of the low declination line concurred with the fixed end point of the low declination line in the other register. For the downward sessions with the rise–fall, the test stimuli with the smallest and the largest excursions are presented in Fig. 2(a) for the LOG versions, in Fig. 2(c) for the ERB versions, and in Fig. 2(e) for the LIN versions. Their corresponding comparison stimuli, i.e., those comparison stimuli that run parallel to them in the corresponding frequency scale, are presented in Fig. 2(b), (d), and (f). Similarly, Fig. 3 shows the test and corresponding comparison stimuli for the upward sessions. Observe that, for each test stimulus, there was a comparison stimulus with a pitch contour running exactly parallel in the corresponding frequency scale. The shifts in the three frequency scales resulted in test stimuli that differed much more from each other than the comparison stimuli, as can be seen in Figs. 2 and 3. An upward shift in semitones resulted in a steeper declination and larger pitch excursions than a shift in Hz. The shift in the ERB-rate scale was somewhere in between. The converse of this was true for the downward shifts.

In anticipation of the results, the pitch contours shown in Figs. 2 and 3 are presented on an ERB-rate scale. For clarity, if subjects perceived prominence of pitch movements on a logarithmic frequency scale, they would match the upper stimulus in Fig. 2(a) to the upper stimulus in Fig. 2(b), and the upper stimulus in Fig. 3(a) to the upper stimulus in Fig. 3(b). However, if they perceived prominence on an ERB-rate scale, they would match these test stimulus to a higher comparison stimulus in the downward sessions, and to a lower comparison stimulus in the upward sessions, as can be discerned from Figs. 2 and 3.

There were nine subjects, all of whom were students or staff members of this institute involved in speech and hearing research. Some were specialists in intonation research, while others were not. Some were musically trained, and one had absolute pitch. None reported hearing defects. Each subject completed all the sessions.

B. Results

There appeared to be no significant difference between the results for the rise, the rise–fall, and the fall. Therefore, the results of these three conditions are collapsed. The average results across all subjects are presented in Fig. 4. The coordinates represent the frequency scale in which the low

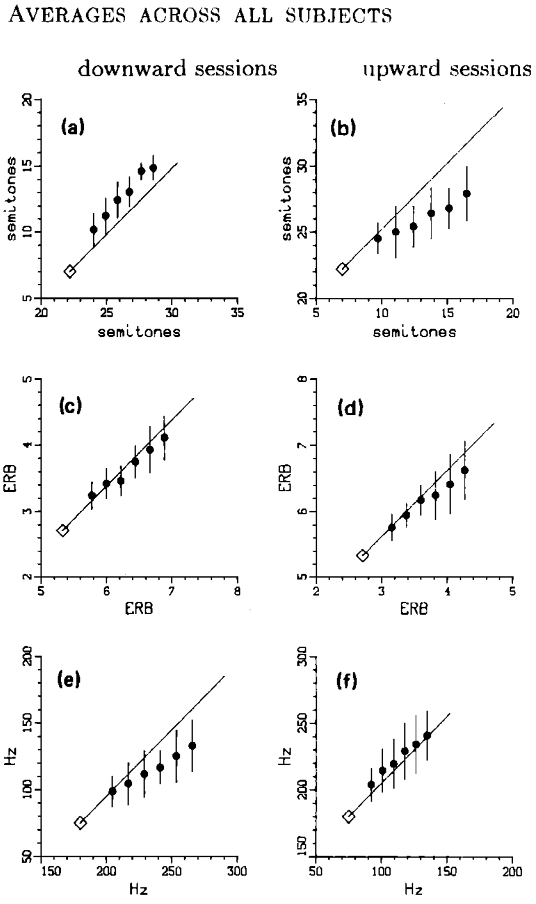

FIG. 4. The average results across all nine subjects who participated in the experiments. The coordinates represent the frequency scale in which the pitches of the low and the high stimulus ran parallel. Thus (a) and (b) shows the results of the sessions with the LOG versions, (c) and (d), those for the ERB versions, and (e) and (f) for the LIN versions. The results for the downward sessions are presented in (a), (c), and (e), while (b), (d), and (f) present the results of the upward sessions. The abscissa of the diamond shows the end point of the lower declination line of the test stimulus, while the ordinate of the diamond shows the end point of the lower declination line of the comparison stimulus. The six different excursions of the test stimuli are presented as the interval between the abscissa of the data points and the abscissa of the diamond. The averages of the matchings by all subjects are presented as the intervals between the ordinate of the data points and the ordinate of the diamond. The vertical bars represent the standard deviation of the results. The straight line with a slope of 45 deg represents the expected outcome if the subject had matched prominence in the frequency scale in which the results are plotted.

and the high stimulus ran parallel. The abscissa of the diamond shows the end point of the lower declination line of the test stimulus (180 Hz for the downward sessions, and 75 Hz for the upward sessions), while the ordinate of the diamond shows the end point of the lower declination line of the comparison stimulus (75 Hz for the downward sessions, and 180 Hz for the upward sessions). The data points represent the averages across all subjects. The abscissa of a data point represents the end point of the upper declination line of the test stimulus, while its ordinate represents the average end point of the upper declination line of the comparison stimulus. This means that the interval between a coordinate of a data point and the corresponding coordinate of the diamond gives the excursion size of the stimulus. The vertical bars represent the standard deviation of the results. The straight line with a slope of 45 deg represents the expected outcome if the subject had matched prominence in the frequency scale in which the results are plotted. In Fig. 4(a) and (b), the results are presented for the LOG versions, in 4(a) for the downward adjustments, and in 4(b) for the upward adjustments. In Fig. 4(c) and (d), the results for the ERB versions are presented, and in Fig. 4(e) and (f), the results for the LIN versions. The results show that for the sessions with the LOG versions, a test stimulus presented in a high register is matched to a comparison stimulus, which, in semitones, has a higher excursion size [Fig. 4(a)]. On the other hand, when the test stimulus is presented in the low register, it is matched to a comparison stimulus which, in semitones, has a lower excursion size [Fig. 4(b)]. For the ERB versions, the test stimuli are matched to a comparison stimulus that has about an equal excursion size in the other register, in this frequency scale. There is a tendency to deviate from the ERB-rate scale for the lowest and the highest excursions of the test stimuli, but it will be shown that this could be attributed to a tendency to match the excursion size of the comparison stimulus to the average of all excursion sizes. For the sessions with the LIN versions, a test stimulus presented in a high register is matched to a comparison stimulus, which, in Hz, has a lower excursion size. When the test stimulus is presented in the low register, there is a tendency to match it to a comparison stimulus which, in Hz, has a higher excursion size.

C. Some comments

Not every subject performed the experiment equally consistently. Some complained about the difficulty of the task. Such subjects produced matchings that tended more to the average of the comparison stimuli, producing a higher variance in the results. These subjects could be selected by comparing their responses in the upward and the downward sessions. If a test stimulus in the low register is matched to a comparison stimulus with some specific excursion size in the high register, a test stimulus in the high register with such an excursion size should, in its turn, be matched to a comparison stimulus in the low register with about the same excursion size as the original test stimulus. This amounts to combining the results of a downward session with those of a corresponding upward session. When the excursion sizes of the lower stimuli are plotted against the excursion sizes of the higher stimuli, there should be a high correlation. This correlation coefficient was calculated for all sessions and for all subjects. This resulted in a quantitative measure that could be used to select subjects responding consistently. All subjects participated in 18 sessions, while for the determination of one such correlation coefficient two sessions were necessary. So, a total of nine correlation coefficients was obtained for each subject. The variances and the bias in the direction of the average were much less when only those five subjects were selected for whom this correlation coefficient exceeded 0.75 in more than six of the nine sessions. Figure 5 shows the average results for the five consistently responding subjects.

After being told that they had matched prominence of pitch movements on an ERB-rate scale, a few subjects felt

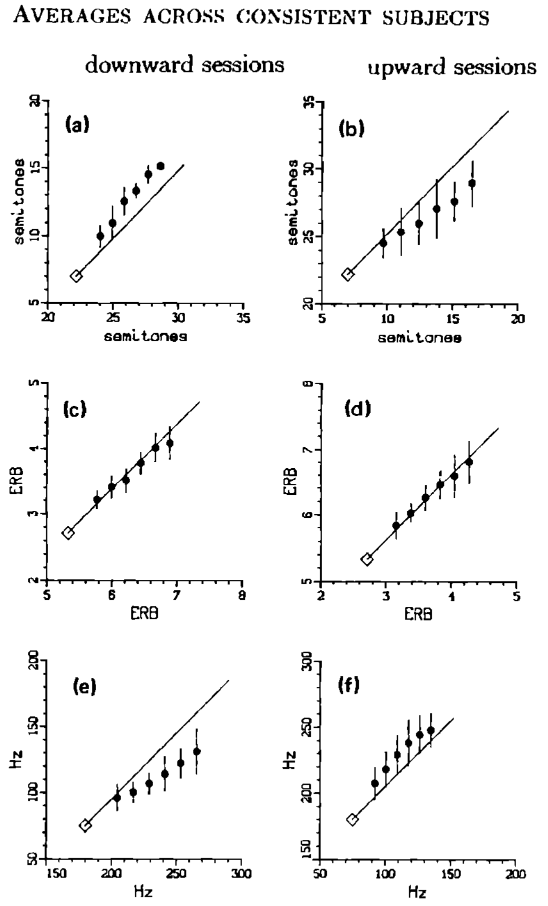

FIG. 5. Average results of the matchings by five subjects showing consistency in their responses; otherwise, as Fig. 4.

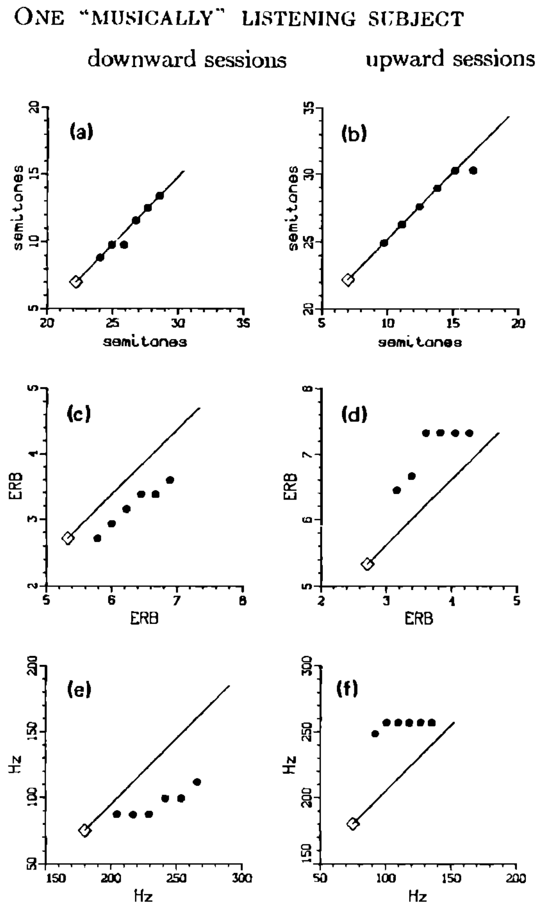

FIG. 6. Results of six sessions with a fall as pitch movement, in which the subject matched the stimuli on a musical scale, and ignored the prominence of the syllables; otherwise, as Fig. 4.

challenged to repeat the experiment, matching the stimuli on a musical, i.e., a logarithmic frequency scale. Some of them succeeded in this, and an example is shown in Fig. 6 for a fall. These subjects reported, however, that this task was very difficult for the kind of stimulus used here. It required a different way of listening, in which the relative pitches of successive syllables had to be analysed in terms of musical intervals between the high and the low declination line. This observation shows that listening musically is a completely different task, in which another perceptual mechanism is used than when prominence of accented syllables is perceived. Incidently, the one subject with absolute pitch also matched pitch movements on an ERB-rate scale.

II. DISCUSSION

The results show that pitch movements in speech intonation can best be expressed in a frequency scale that is derived from the frequency selectivity of the auditory system. Excursions that are equal when expressed in Hz or in semitones do not have the same prominence when presented in different registers. These results do not say anything about what in a pitch contour lends prominence to a syllable, e.g., the excursion size, or the slope of the pitch movements. Since the traditional critical-band scale with the Bark as unit is linear under 500 Hz, these results show that the ERB-rate scale is to be preferred as far as speech intonation is concerned.

The experiment included three different prominence-lending pitch movements. For all cases, the same conclusion could be drawn. These three pitch movements lasted a relatively short time and extended over not much more than one syllable. The experiment did not include slower pitch movements that extend over various syllables, as occur in Dutch. Declination was also included, but not varied as an independent variable. So, theoretically, it remains possible that declination and slow pitch movements are perceived on another scale. Nothing was observed that supported this idea, however. Therefore, it is concluded that the number of critical bands crossed by a pitch movement, or the velocity with which critical bands are crossed, determines the extent to which a pitch movements contributes to the prominence of a syllable.

There are some data in the psychoacoustic literature which corroborate these results, when speech is considered as a frequency-modulated (FM) sound signal. Based on Fechner's hypothesis that magnitude perception can be obtained by integration of just-noticable-differences (jnd's) [see also Suchowerskyj (1977) and Houtsma *et al.* (1980)], Zwislocki (1965, p. 49) derived a nearly exact Mel scale from jnd measurements for FM sine waves. On the other hand, Moore and Glasberg (1989) found that jnd's still differed by a factor of 2, when expressed as fractions of ERBs.

They compared these jnd's for FM within one tone with jnd's for frequency discrimination of two successive tones of constant frequency, and concluded that different mechanisms may account for the difference between these jnd's. This may also hint at different mechanisms underlying the perception of pitch in intonation and the perception of musical melodies, which, there is no doubt, are perceived as identical when they have pitches that run parallel on a logarithmic frequency scale. Klatt (1973) has measured jnd's for pitch discrimination and 't Hart (1981) for pitch distance discrimination in sounds which "included the dynamic qualities characteristic of speech" (Klatt, 1973, p. 8). Their data are not accurate enough, however, to draw any conclusions concerning the scale on which these jnd's are constant.

Also in vowel perception, there are some discussions on whether formant frequency is perceived on a logarithmic or a scale derived from the frequency selectivity of the auditory system (e.g., Nearey, 1989; Miller, 1989). No conclusive experiments have been reported, however. This is probably partly due to the fact that no operational experimental paradigm is known in which equality in different frequency regions can be established.

From the result obtained in this study, some conclusions can be drawn on the way in which pitch is coded in the central nervous system. In speech intonation, the prominence that a pitch movement lends to a syllable appears to be well defined perceptually, so that excursion sizes in different frequency regions can be compared. This was used to determine on which frequency scale pitch movements in speech intonation are judged equal. A frequency scale derived from the frequency selectivity of the auditory system fitted the results best. Since, in speech, most harmonics have frequencies higher than 500 Hz, and also the ERB-scale is nearly logarithmic above 500 Hz, prominence of pitch movements would be perceived on an approximately logarithmic frequency scale, if perceived prominence were based on a combination of the excursions of the harmonics. Since this appears not to be the case, it must be concluded that perceived prominence is based on the course of pitch itself and not of its harmonics. This means that there is a pitch-coding array in the human speech processor. It has now been shown that this array has the same linear organization as the array of filters in the peripheral auditory system.

ACKNOWLEDGMENTS

This work was supported by the Instituut voor Doven, Sint-Michielsgestel. We are grateful to Hans 't Hart and Jacques Terken for their fruitful discussions and constructive comments on the manuscript.

Cooper, W. E., and Sorensen, J. M. (**1981**). *Fundamental Frequency in Sentence Production* (Springer-Verlag, New York).

Fletcher, H. (**1953**). *Speech and Hearing in Communication* (Van Nostrand, New York), pp. 153–175.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Modern Phys. **12**, 47–65.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Graddol, D. (**1986**). "Discourse specific pitch behavior," in *Intonation in Discourse*, edited by C. Johns-Lewis (Croom Helm, London), p. 221–237.

Greenwood, D. D. (**1961**). "Critical bandwidth and the frequency coordinates of the basilar membrane," J. Acoust. Soc. Am. **33**, 1344–1356.

Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.

Hamon, C., Moulines, E., and Charpentier, F. (**1989**). "A diphone synthesis system based on time-domain prosodic modifications of speech," Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP-89, pp. 238–241.

Houtsma, A. J. M., Durlach, N. I., and Braida, L. D. (**1980**). "Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching," J. Acoust. Soc. Am. **68**, 807–813.

Klatt, D. H. (**1973**). "Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception," J. Acoust. Soc. Am. **53**, 8–16.

Liberman, M. C. (**1982**). "The cochlear frequency map for the cat: Labeling auditory-nerve fibers of known characteristic frequency," J. Acoust. Soc. Am. **72**, 1441–1449.

Miller, J. D. (**1989**). "Auditory-perceptual interpretation of the vowel," J. Acoust. Soc. Am. **85**, 2114–2134.

Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J., and Glasberg, B. R. (**1986**). "The role of frequency selectivity in the perception of loudness, pitch and time," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 251–308.

Moore, B. C. J., and Glasberg, B. R. (**1989**). "Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation," J. Acoust. Soc. Am. **86**, 1722–1732.

Neary, T. M. (**1989**). "Static, dynamic, and relational properties in vowel perception," J. Acoust. Soc. Am. **85**, 2088–2113.

Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Acoust. Soc. Am. **59**, 640–654.

Rietveld, A. C. M., and Gussenhoven, C. (**1985**). "On the relation between pitch excursion size and prominence," J. Phon. **13**, 299–308.

Stevens, S. S., Volkman, J., and Newman, E. B. (**1937**). "A scale for the measurement of the psychological magnitude pitch," J. Acoust. Soc. Am. **8**, 185–190.

Stevens, S. S., and Volkman, J. (**1940**). "The relation to pitch and frequency: a revised scale," Am. J. Psychol. **53**, 329–353.

't Hart, J. (**1981**). "Differential sensitivity to pitch distance, particularly in speech," J. Acoust. Soc. Am. **69**, 811–821.

Suchwerskyj, W. von (**1977**). "Beurteilung von Unterschieden zwischen aufeinander folgenden Schallen," Acustica **38**, 131–139.

't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation* (Cambridge U. P., Cambridge, England).

Traunmüller, H., Branderud, P., and Bigestans, A. (**1989**). "Paralinguistic speech signal transformations," Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS) **10**, 47–64.

Traunmüller, H. (**1990**). "Analytical expressions for the tonotopic sensory scale," J. Acoust. Soc. Am. **88**, 97–100.

Wilson, J. P., and Evans, E. F. (**1977**). "Cochlear frequency map for the cat," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, New York), p. 69.

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). "Critical band width in loudness summation," J. Acoust. Soc. Am. **29**, 548–557.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., and Terhardt, E. (**1980**). "Analytical expressions for critical-band rate and critical bandwidth as a function of frequency," J. Acoust. Soc. Am. **68**, 1523–1525.

Zwislocki, J. (**1965**). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology, Vol. III*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), pp. 1–97.