

Differential sensitivity to pitch distance, particularly in speech

Citation for published version (APA):

Hart, 't, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America*, 69(3), 811-821.

Document status and date:

Published: 01/01/1981

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.

Differential sensitivity to pitch distance, particularly in speech

Johan 't Hart

Institute for Perception Research, P.O. Box 513-5600 MB Eindhoven, The Netherlands
(Received 5 October 1978; accepted for publication 7 November 1980)

The fundamental frequency in speech shows many rapid variations, part of which determine the perceived shape of the pitch contour. This implies that the accuracy with which listeners perceive changes of F_0 is more relevant to understanding the perception of intonation than the traditional just noticeable difference of F_0 in speech. This study examines the sensitivity to differences in the amount of change of F_0 , upward (Experiment Ia) and downward (Experiment Ib). Subjects, 74 and 104, respectively, with widely different musical ability can be divided into three categories: (1) Quite a number of them were not able to discriminate differences of less than 4 semitones (nondiscriminators); (2) other subjects wrongly tried to base their judgments on a simple comparison of the final pitches of a stimulus pair (final pitch discriminators); (3) the remaining subjects (pitch distance discriminators) yielded average jnd's of about 1.5 to 2 semitones. Since the issue is associated with musical interval sense, similar experiments were carried out using piano tones. The results were essentially the same as with the speech stimuli. The outcome suggests that only differences of more than 3 semitones play a part in communicative situations.

PACS numbers: 43.70.Dn, 43.66.Hg

I. INTRODUCTION

A. Pitch judgment accuracy

The psychoacoustic literature of the last decades leads one to suppose that relatively little remains to be learned about the accuracy with which pitch can be perceived under various conditions. But between a number of these conditions considerable discrepancies in accuracy have been found and it is not immediately clear why these exist.

Let us first consider the situation in which subjects have to tell the pitch of one single pure tone. Only a very small percentage of individuals appear to have an accuracy of $\frac{1}{4}$ to $\frac{1}{5}$ of a semitone. The majority have an average error of 5 to 9 semitones (Bachem, 1937) [this may well correspond to Pollack's (1952) perfect identification of only 5 to 7 tones in the range from 100-8000 Hz]. Persons of the former type are said to possess absolute pitch, which is, of course, something different from understanding the mechanism behind their remarkable performance.

Upon introducing a second tone, we may examine the just noticeable difference (jnd) in frequency. Surprisingly enough, we find that people who are entirely unable to say whether one single given tone has a frequency of 800, 1000, or 1200 Hz, may very well be able to distinguish a tone with a frequency of 1000 Hz from one with 1005 Hz (the jnd for best performers at 1 kHz is well below 1 Hz, e.g., Ritsma, 1965; Nordmark, 1968; Rakowski, 1971). It is not unknown, however, that listeners can be found who initially, at least, require much larger differences in frequency to discern two tones as different in pitch.

Quite a different task confronts the observer when he is asked to identify the musical interval separating two simultaneously presented tones. It is not surprising that the accuracy is much less: Plomp *et al.* (1973),

for example, have found a "standard deviation" of about $\frac{1}{4}$ semitone, with musically trained subjects.

This difference in accuracy demonstrates that estimating the distance between two tones is much more difficult than establishing that such a distance is different from zero. This effect is also illustrated in experiments in which, with three tones, the subjects are asked to adjust the frequency of the second tone in such a way that the distance between a given tone No. 1 and tone No. 2 is judged equal to that between tone No. 2 and a given tone No. 3: Stevens and Volkman (1940) have found average intra-observer variabilities of 10%-15% in experiments of this kind.

It seems that the ability to estimate the interval between two tones reliably rests on an "absolute judgment of the distance of tones" rather than on the "possession of relative pitch" as this ability is usually referred to. The essence of this ability is, reportedly (e.g., Révész, 1946), to be able almost infallibly to transpose given musical intervals. From that point of view, the Stevens and Volkman setup is only a special case of a more general four-tone situation. With four tones, the distance between Nos. 1 and 2 have to be compared to that between Nos. 3 and 4. Experiments of this kind were long ago described by Stumpf (1890, Vol. I, p. 247, Vol. II, p. 403: "Urteile über Tondistanzen"), and Révész incorporated this setup in his tests of musical talent. More recently, Attneave and Olson (1971) measured this ability in two musical and four nonmusical subjects by asking them to transpose patterns built of two tones to a frequency region which was always an integral number of octaves minus half an octave away, and in which one of the tones was given as an anchor and the second tone had to be adjusted. They were not primarily interested in the accuracy as such, but they mention 2% or less variability (below 5 kHz) for the musical subjects, whereas for the nonmusical subjects,

it can be seen to increase from 2%–6% for small intervals up to 8%–18% for the octave. Burns (1974) tried to determine whether performers of classic Hindustani music, in which the octave is divided in 22 steps instead of 12, do indeed possess a more refined interval sense. Using four tones, none of which were equal, he asked his subjects whether they judged the interval between tones 1 and 2 larger than, equal to, or smaller than that between tones 3 and 4. Only one of the 13 subjects was able with any accuracy to say which of the two given intervals was the larger.

B. Speech pitch judgment accuracy

The central problem of the present study is, how detailed a description of F_0 variation in speech should be in order to be sufficiently precise in view of the perceptual limitations of the human listener. Basing herself on the outcome of Flanagan's and Saslow's (1958) measurement of pitch discrimination for synthetic vowels, which is 0.3% to 0.5%, Lehiste (1970) opts for an accuracy of 1 Hz. However, Klatt (1973) questions the relevance of Flanagan's and Saslow's outcome, since the experimental conditions did not include the dynamic qualities characteristic of speech.

Meanwhile, jnd's for F_0 have also been reported for static conditions with speech signals. For example, Isačenko and Schädlich (1970) have found 3% (of 150 Hz), and Rossi and Chafcouloff (1972) 4% (of 195 Hz).

Studies of dynamic pitch phenomena have produced uncomfortably large discrepancies: Klatt's (1973) subjects could distinguish downward glides from upward glides in 250-ms synthetic vowels if the difference in F_0 between beginning and end of each glide was only 1.5 Hz; Rossi (1971), however, reports that in the same amount of time an upward movement of 19 Hz can only just be heard as being a "glissando" at all. Witting (1962), in an earlier attempt to reconcile the discrepancies between the objective course of F_0 (as measured by means of a Lottermoser–Grützmaier F_0 meter) and the transcriptions of perceived pitch phenomena in speech, has given operational rules for the smoothing effect caused by the limited temporal resolution and pitch discrimination in dynamic speech situations. Basing his calculations and estimations on the Shower and Biddulph data (1931) rather than on direct psychoacoustic measurements of sensitivity to dynamic aspects of pitch, he proposes that changes of F_0 in voiced segments shorter than 50 ms should be averaged, and changes of less than 5 Hz should be leveled out in segments longer than 50 ms.

Likewise, Di Cristo (forthcoming) has developed a transformation system for reshaping F_0 measurements into perceived contours, applying the various thresholds that have been established by Rossi *c. s.*

II. DISCRIMINABILITY OF THE AMOUNT OF CHANGE OF F_0 IN SPEECH AND OF MUSICAL INTERVALS BETWEEN PIANO TONES

A. Background considerations

In this study we will confine ourselves to examining the sensitivity to differences in the amount of change of

F_0 , rather than in the rate of change; we will thus deal with only one aspect of the entire complex of the perceptual abilities with respect to dynamic pitch phenomena.

The main reason for this choice is the following: In our earlier work on intonation of Dutch ('t Hart and Cohen, 1973; 't Hart and Collier, 1975), we have given descriptions of the properties of stylized pitch contours that were demonstrated to be "perceptual equivalents" of actual courses of F_0 in natural speech. In the stylized pitch contours, use was made of changes of F_0 , or "pitch movements," with standard specifications as to slope and duration and hence the amount of change of F_0 , or the size of the pitch movement. However, a comparison of stylized and natural intonation does of course show considerable discrepancies, notably with respect to these sizes. With a view to a possible future adequate account of the variability of the sizes of pitch movements in natural speech, we would like to know the degree of refinement necessary to cover this variability, preferably on the basis of the perceptual abilities. Moreover, insofar as pitch movements may lend prominence, one could ask, in the case of an utterance in which two syllables are made prominent by means of similar pitch movements, how much larger either of the two pitch movements should be in order to make that syllable more prominent than the other.

We decided to try to measure the discriminability of the size of pitch movements and of musical intervals by means of a paradigm similar to the four-tone situation. That is, in the case of speech stimuli, the sizes of two pitch movements had to be compared; in the case of piano tone stimuli, the interval between tone 1 and 2 had to be compared with that between tone 3 and 4. We were primarily interested in the speech experiments; the piano tone interval experiments do provide an intermediate step between the domain of pitch discrimination between two stationary complex tones and that of the magnitude discrimination between two pitch movements.

B. Procedure

1. Experiment 1: Pitch movements in speech

Using a somewhat adapted, forced-choice pair comparison procedure, subjects were presented on each trial with a pair of stimuli. Each stimulus contained a pitch movement of variable size; both movements in a pair were in the same direction. The subject's task was to indicate which of the two members contained the larger movement. Rising and falling movements were used in separate blocks of trials and therefore constituted separate experiments.

The range of pitch distances covered that of normal, nonemphatic speech. In such conditions, the size of prominence lending rises and falls rarely exceeds half an octave (in Dutch intonation). Therefore it was decided to use a range from 1 to 6 semitones. This gave the additional advantage that the musical fifth was avoided, which, according to Plomp *et al.* (1973), is very easily confused with the musical fourth.

Since it was our aim to measure the discriminability of the size of pitch movements or jumps rather than that of pitch proper, the stimulus pairs contained no fixed reference; i. e., in 67% of the cases, both the initial and the final F_0 were different in the two members of a stimulus pair. In 33% of the cases, either the initial or the final frequencies were equal. Thus the stimulus material was fairly representative of the various situations that are observed in actual speech. In Dutch intonation, no two successive prominence-lending rises occur without the insertion of a "reset" to the base line, the so-called nonfinal fall. Similarly, there is a rise between almost any pair of successive falls. But as a result of the declination (the slowly running-down frequency all over an utterance), any such movement later in an utterance starts at a lower frequency than an earlier otherwise comparable movement. How high they end depends on their relative sizes. As soon as a new declination line is initiated, the next movement will start higher than the previous one. Finally, two utterances of approximately the same duration tend to start at the same height. Meanwhile, since the declination line is only slightly tilted, the observed differences are not very large. In the stimulus material, differences between initial frequencies (tone 1 and tone 3) were therefore limited to plus or minus 3 semitones and plus or minus 6 semitones, of which the former occurred three times more often.

2. Experiment II: Musical intervals on the piano

The conditions of the musical interval experiment were similar to those of the speech experiment. A stimulus pair contained four tones, of which the two intervals between Nos. 1 and 2 and between Nos. 3 and 4 were the ones that had to be compared by the listeners. Again, in the majority of the stimulus pairs, both the initial and the final tones were different, so that none of the four tones were equal. In other pairs, either tone 1 was equal to tone 3, or tone 2 was equal to tone 4.

C. Stimulus specification

1. Experiment Ia: Speech rises

Figure 1(a) provides an example of one of the stimulus pairs of experiment Ia. They were obtained in the following way: Six different four-syllable Dutch number names were processed by means of the Intonator (Willem, 1966) so as to provide them with pitch contours that may actually occur in short utterances of Dutch (and particularly in the enumeration of number names), with slowly running down pitch (the declination line) during all syllables except for the third, accentuated one, which had a rise of 1, 2, ..., up to 6 semitones.

The variation in the size of the movements should not be obtained by means of varying their slope, since that might constitute an extra cue for the listeners. Variation of the duration—which was actually applied—could give an extra cue as well, but this was avoided by having the rises start and end in a voiceless prevocalic consonant. Similarly, in experiment Ib, the falls started at a fixed point in the voiced part of the accented syllable (50 ms after VOT) and ended in a voice-

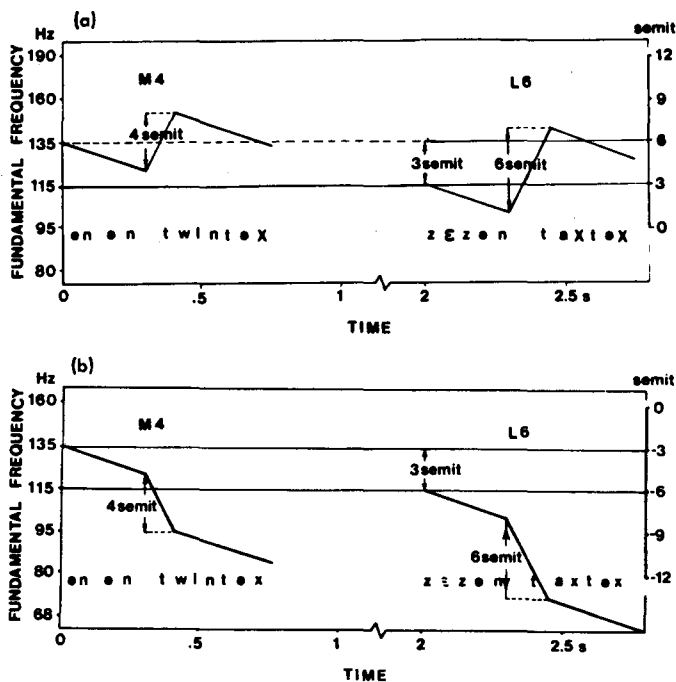


FIG. 1. (a) Example of a stimulus pair (two intonator processed Dutch number names) of experiment Ia. The first stimulus, M4, is situated in the midregister and contains a rise of 4 semitones; the second stimulus, L6, with a rise of 6 semitones in the low register should be recognized as having the larger pitch movement. (b) Example of a stimulus pair of experiment Ib.

less postvocalic consonant. The voiced part of the accented syllable should be short so as to ensure timely ending of voicing in the case of a 1-semitone fall, where the duration of the movement was only 25 ms. These conditions were met by taking the following number names: /enən'twIntəx/, /ɑχtən'twIntəx/, /twejən'sestəx/, /veivən'sestəx/, /drijən'təxtəx/, /zezən'təxtəx/. The initial frequency of a contour was taken as 115, 135, or 160 Hz, thus situating the contour in a low, mid, or high "register," as we will call them for the sake of convenience. The registers were approximately 3 semitones apart. The choice of the particular number name was arbitrary with respect to the register and to the size of the pitch movement. Thus there were actually eighteen different stimuli.

Since any of the eighteen contours obtained could serve as both "reference" and "variable" signal in the proposed procedure, 324 stimulus pairs were possible. From these, a selection of 90 was made in such a way that smaller differences in the size of the pitch movements occurred up to three times as frequently as did the larger. There were twelve stimulus pairs with equal sizes in the two stimuli, and these were always in different registers. There were as many pairs in which the first item had the larger excursion as inversely, and these were evenly distributed among the different register situations.

2. Experiment Ib: Speech falls

The same number names were now provided with falling pitch movements instead of the rises of experiment Ia, likewise with sizes between 1 and 6 semitones, in the same three registers [see Fig. 1(b)]. A similar,

but somewhat different selection of 100 pairs served as stimulus material in experiment Ib.

3. Experiment II: Piano tone intervals

For the second experiment, 100 stimulus pairs were played on the piano to provide an analog of experiments Ia and Ib. A stimulus pair consisted of four tones, of which Nos. 1 and 3 again were either low, mid, or high, Nos. 2 and 4 were 1–6 semitones higher in experiment IIa, lower in experiment IIb. The initial frequencies of experiments Ia and Ib were best approximated by taking B flat, D flat, and E. Thus, for example, in experiment IIa, with rising intervals, stimulus pairs Nos. 27 and 28 were as follows:



The pianist synchronized the tones with the temporal organization of the stimuli of experiments Ia and Ib by means of 1-s interval impulses, which were played back from the second track of the tape on which the tones were recorded, and made audible through headphones. The one-hundred tone quadruplets had the same specifications as those of experiment Ib, with inverse direction of the intervals.

Experiment IIb, with falling piano tone intervals, had the same stimuli as experiment IIa, but with downward-going intervals. Stimulus pairs Nos. 27 and 28 were now:



Figure 2 shows the temporal organization of the test tapes. The warning signal, presented 2.5 s prior to each stimulus pair, was a 1-kHz pure-tone burst in the speech experiments, a staccato *c* in the piano tone experiments. In the latter, moreover, the blocks of Fig. 2, which represent the number names in experiments Ia and Ib, should be read as two portato tones each, at 0.5-s time interval; the warning signal came 2 s prior to the first tone.

D. Subjects, instruction, and conditions

In an attempt to obtain a representative sample of the arbitrary (Dutch speaking) listener, subjects were primarily taken among people not experienced in taking part in psychoacoustic experiments. In all experiments together, 16% of them were psychoacoustically untrained colleagues from the laboratory, and the majori-

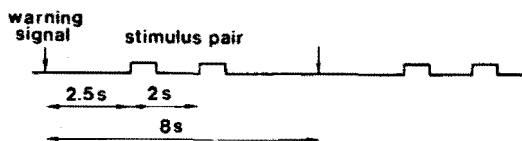


FIG. 2. Temporal organization of the test tapes. A short 1-kHz pure tone burst precedes the stimulus pair. Duration of each of the stimuli: 750 ms. The subjects are instructed to respond prior to the warning signal of the next stimulus pair.

ty, 68%, were outsiders, partly students (of medicine and psychology), partly arbitrary. Ages varied from 15–55, both males and females were present, with no obvious hearing impairments. About half of the subjects from the laboratory took part in all four experiments, the other half only in Ib and IIa. The “other subjects” of Ia and Ib were the same individuals, those of IIa were different from those of IIb.

It was desirable to collect additional data from psychoacoustically trained subjects. So, again for all experiments together, 16% of the subjects belonged to either those working in psychoacoustics or in experimental phonetics in the laboratory. Most of the laboratory subjects did the experiments twice, other subjects only once. For instance, experiment Ia was done twice by 15 laboratory subjects and once by 44 other subjects, thus yielding a total of 74 runs. The subjects from the laboratory listened through headphones in individual sessions; the other subjects did the experiments collectively, listening to loudspeakers.

In view of the aim of the experiment, we deliberately refrained from training the subjects to some considerable extent, apart from what was necessary to make them familiar with their task. On the one hand, we were interested in the everyday-life performance of arbitrary native users of the language rather than in the utmost that can be achieved in the laboratory with highly trained, talented individuals. (From that point of view, the use of headphones—chosen for mere practical purposes—is already somewhat artificial.)

On the other hand, native users of the language are fully trained to hear the features that are perceptually relevant so that, if care has been taken to make the stimuli sufficiently similar to what listeners are exposed to in normal speech, extra training is superfluous. This should then be reflected in the data from subjects who took the task more than once: They should not perform better in the second run.

The subjects were informed about the nature of the stimuli they were going to hear. They were instructed to score on their answer sheets in either of two columns: Left, if the first item of the stimulus pair was judged to contain the larger pitch movement or interval, right, if they thought it was the case with the second item. Since it might constitute a problem for some subjects that they are not accustomed to judge aspects of the speech signal in isolation and/or in its quality as such, the instruction said that the larger pitch movement might be heard as a stronger accent. Subjects were not forced to respond on each trial, but rather strongly encouraged to guess in cases where they could not hear any difference. If they felt entirely unable to guess, or entirely convinced that the movements (intervals) were equally large, they were allowed to score with the sign “=” In such cases, they were told, the score would be selected by lot. Extra care was taken with respect to the instruction of the subjects from outside the laboratory: An instruction tape contained several examples, and a warning against a certain type of stimulus pair, of which the one in Fig. 1(a) is an example, where the first item has the highest pitch peak,

but the second one the larger pitch movement.

Subjects who might possess absolute pitch were warned that, due to a difference in tape speed of about 3% between the machine on which the piano tones had been recorded and the one on which they were played back, they should rely on their interval sense rather than on their absolute pitch. The time between the presentation of two pairs of tones was sufficient to allow musically trained subjects to identify each interval; they were told that they were allowed to "calculate" their judgment about the relative size of the intervals on that basis. It is important to note that this strategy can only be applied fruitfully in the piano tone experiments, since the estimation of musical intervals in speech pitch movements is highly unreliable. For both types of stimuli, a number of listeners found it helpful to use kinesthetic means like singing or whistling. Subjects who participated in collective sessions were not encouraged to do so.

III. RESULTS

In a preliminary analysis of the data, percent correct answers were plotted as a function of the absolute value of the difference between the sizes of the movements (intervals) for all subjects individually. This procedure provides a smoothing of the otherwise rather irregular curves, and a rough measure of sensitivity, using a 75% correct response criterion. For a considerable number of subjects, particularly in experiments Ib and IIb, the curves remained far below the 75% criterion; the sensitivity threshold of these subjects is taken to be more than 5 semitones. Apart from this remarkably low sensitivity in some subjects, it was perhaps more surprising that also, in general, the performance was rather poor. Figure 3 shows the histograms of the sensitivity threshold for the four experiments (the hatched parts correspond to the untrained subjects). They appear to be bimodal, with a boundary at about 4 semitones. Therefore the subjects were divided into two classes: "Nondiscriminators" with thresholds of 4 semitones or more, and "discriminators" who needed less than 4 semitones. But the mean thresholds for the discriminators are not less than 2.1-2.8 semitones. That is much higher than would be expected on the grounds of known data about the just noticeable difference in pitch of speech sounds, even with untrained subjects.

Although the outsiders had a higher percentage of nondiscriminators in all four experiments, there was no significant difference between the performance of the discriminators among the outsiders and those among the (untrained) laboratory subjects. This suggests that neither their being unacquainted with psychophysical experiments, nor the use of a loudspeaker in collective sessions is the primary cause of the high values of the mean thresholds.

At this point, it seems interesting to consider the results of the psychoacoustically trained subjects. These people are experienced in pitch perception experiments, but they were not trained beforehand in this particular experiment. Yet, their performance was significantly

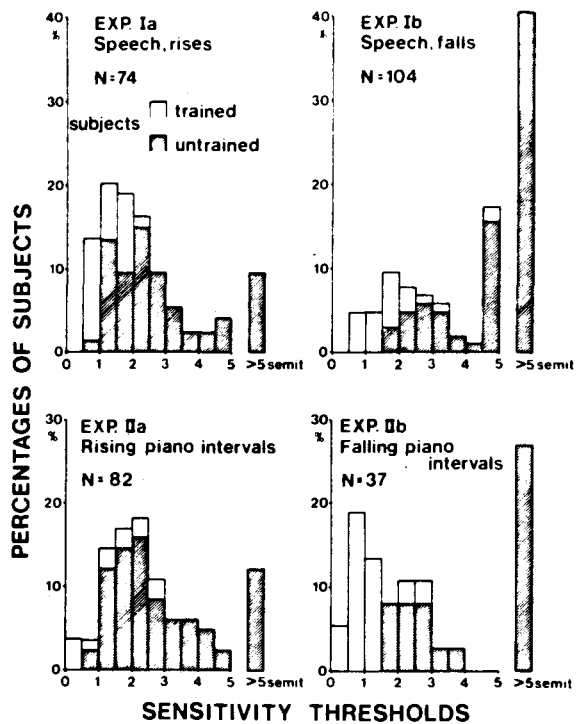


FIG. 3. Histograms of the 75% correct response sensitivity thresholds in the four experiments.

better than that of the other subjects. Their results are represented by the open bars of Fig. 3. This group of subjects contained no nondiscriminators, with one exception, in experiment Ib.

It would have been possible to train this group further to see if their performance would still ameliorate. This was not done, for two reasons. The main reason was that, in view of our central problem, such a result would not be representative of the situation in which an arbitrary user of the language listens to normal speech; the second reason was that all subjects who did the experiments twice did not perform better in the second run. This is true for trained and untrained subjects separately. It was decided not to keep trained and untrained subjects apart in further analysis for several reasons. First of all, the significantly different performance of trained and untrained subjects is largely due to the "extremely low thresholds" (still higher than 0.5 semitone, or 3%) in a small number of trained subjects, who were the same individuals throughout the four experiments. Moreover, especially in experiments Ia and IIa, some untrained subjects also had very low thresholds. Finally, adding the data from these fine discriminators to those from the others will only flatter the total outcome, to the effect that our final estimate of the required accuracy will be at the safe side. Further analysis will be restricted to the data from the discriminators. One last remark about nondiscriminators: With both speech falls and piano falls, their percentages are larger than with corresponding rises. Since, especially in the speech experiments, the majority of the subjects were the same individuals in both experiments, it may be deduced that, at least in

speech stimuli, falls are more difficult to judge than rises. For instance, there were no discriminators with speech falls who were nondiscriminators with speech rises.

A. Speech pitch movements

We now turn to a closer analysis of the speech experiments in the first place. An examination of the incorrect responses by all subjects together, nondiscriminators included, revealed that, with very few exceptions, all stimulus pairs with more than the average number of incorrect responses shared the following property: The extreme pitch (highest in the case of rises, lowest in the case of falls) occurs in one member of the stimulus pair, but the largest movement or interval occurs in the other [see Fig. 1(a)].

In experiment Ia, the average number of incorrect responses per stimulus pair of this class was 41 and only 13 with the other stimulus pairs. This can be explained on the basis of the supposition that in many instances subjects were not able to recall the initial pitch of the first stimulus, so that the only basis for their judgment was the difference in final pitches. This is fatal with the class of stimulus pairs at issue, which we will call "trick stimulus pairs."

Since subjects often compared final pitches instead of sizes of movements, it is possible that some subjects always performed the task in this way. It seemed justified to try to trace those subjects who did not follow the instruction and to analyze their performance in a different way. Therefore, for each subject individually, the responses to trick stimulus pairs were compared with those to stimulus pairs of a similar degree of difficulty, and a chi-square test was applied to see if there were significantly more incorrect responses to trick stimulus pairs than to the comparable pairs. The chi-square values obtained constituted a continuum. The necessary cut was made at a level of significance with $p < 0.001$. The procedure selects final-pitch discriminators irrespective of their individual sensitivity thresholds: If the few incorrect responses of a good performer are mainly given to trick stimulus pairs, he may just as well be characterized as a final pitch discriminator. Accordingly, thresholds for final pitch discriminators were found all over the range from 1 up to more than 5 semitones; on the other hand, the pro-

portion of final pitch discriminators among those with thresholds less than 1 semitone was extremely small.

Table I shows, for both experiments, the total number of "subjects" (in which two "subjects" may represent two runs by one individual), numbers and percentages of discriminators and nondiscriminators and among the former category, of final-pitch discriminators, and finally, of the subjects who are taken to have based their judgments on a comparison of the pitch distances in either member of a stimulus pair. The results of the pitch distance discriminators were analyzed by computing the percentages of answers: "Second movement larger" (right column) as a function of the size of the second movement minus the size of the first one.

It seemed justified to introduce an average subject by summing these percentages and dividing them by the number of subjects. This has been done for the entire stimulus material (of each experiment), and also after having it split out in stimulus pairs with equal register, register up (initial pitch of second stimulus 3 or 6 semitones higher than that of first one), and register down. The results are shown in Figs. 4 and 5. The dots are the means of percentages, the curves are best fits by means of the function $P(x) = 100 / \{1 + \exp[-1.1(x - x_0)/d]\}$, in which d is the jnd and $P(x_0) = 50\%$. Half the distance between the 25% and 75% points was taken as a measure of the just noticeable difference; the 50% intersection gives the point of subjective equality.

Figures 4 and 5 show that the just noticeable difference is typically less in the equal register condition than in the conditions with shifting registers. The point of subjective equality is close to zero for "overall" and "registers equal" indicating that a potential time-order bias is negligible, but it has considerable deviations in shifting register conditions: "Second movement larger" answers are favored with rises and register up (i. e., the second stimulus of a pair starts higher), and with falls and register down; they are suppressed with rises and register down, and with falls and register up. Table II gives a survey of the results.

The scores of the 17 final pitch discriminators of experiment Ia and the six of experiment Ib have primarily been analyzed in the same way. See Table III.

With registers equal, the jnd's do not differ from

TABLE I. Categorization of subjects. Indicated are those who could not do the discrimination, those who applied the wrong strategy, and those who survived these selection criteria.

Experiment	Ia: Speech rises	Ib: Speech falls
Total number of subjects	74	104
Nondiscriminators	12 (16%)	63 (61%)
Discriminators	62 (84%)	41 (39%)
Final pitch discriminators	17 (27%)	6 (15%)
Discriminators: Pitch distance discriminators	45 (73%)	35 (85%)

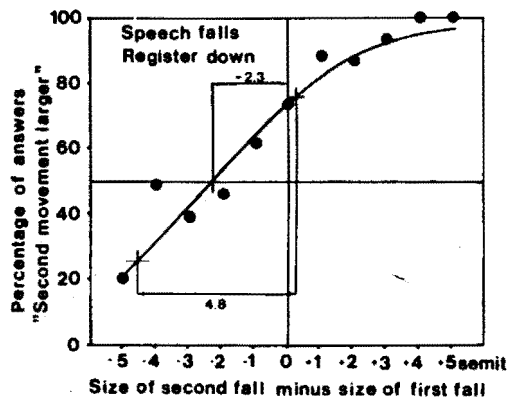
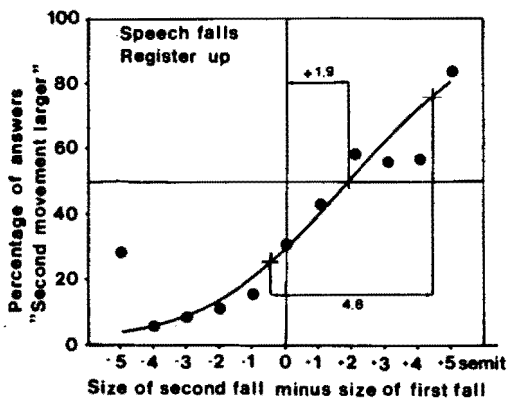
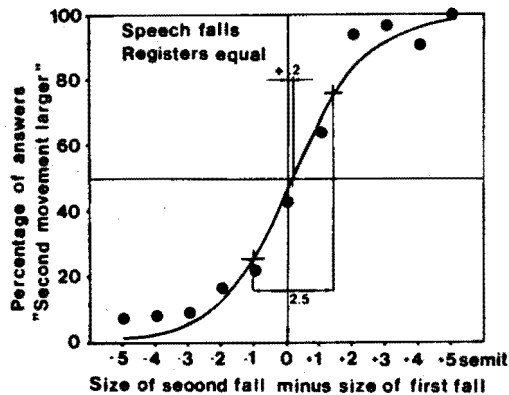
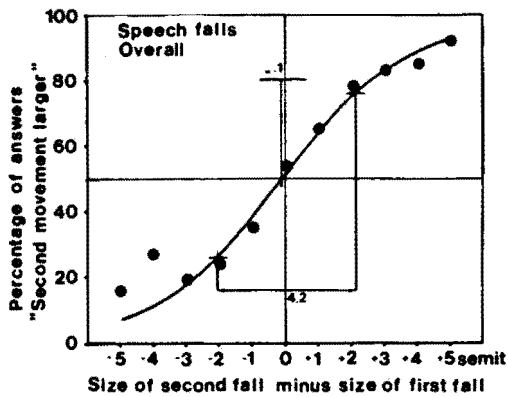
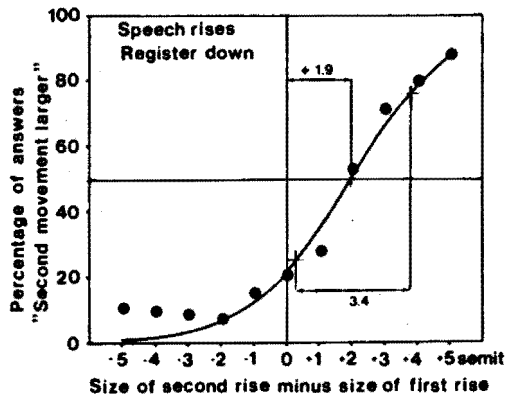
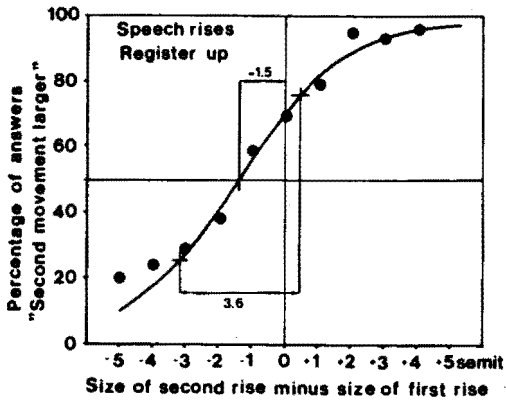
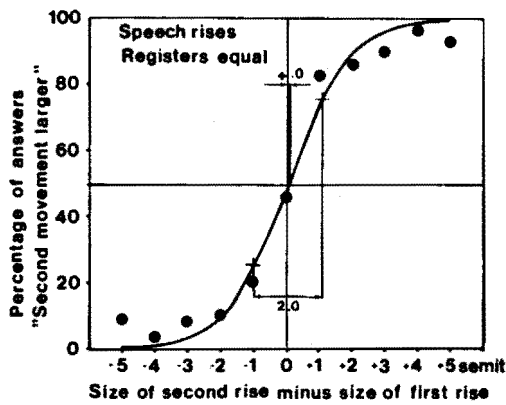
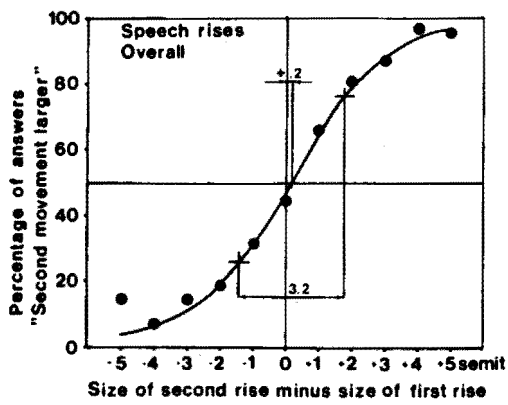


FIG. 4. Psychometric curves for the average pitch distance discriminator in experiment Ia: Speech rises; separated for (a) the entire material, (b) stimulus pairs in equal registers, (c) those with register up, in which the second stimulus starts at a higher frequency, and (d) those with register down. Indicated are the 25% and 75% intersections, giving twice the jnd, and the 50% intersections, showing the points of subjective equality.

FIG. 5. Psychometric curves for the average pitch distance discriminator in experiment Ib: Speech falls; as in Fig. 4.

TABLE II. Just noticeable differences and points of subjective equality (pse) in semitones for the average pitch distance discriminator. The data are given for the stimulus material as a whole, and for each of the three possible register situations (e.g., register up: Second stimulus starts 3 or 6 semitones higher than first one).

Experiment Number of subjects	Ia: Speech rises 45		Ib: Speech falls 35	
	jnd	pse	jnd	pse
Overall	1.6	+0.2	2.1	-0.1
Registers equal	1.0	+0.0	1.3	+0.2
Register up	1.8	-1.5	2.4	+1.9
Register down	1.7	+1.9	2.4	-2.3

those of the pitch distance discriminators and the points of subjective equality are again close to zero. In the shifting register conditions, the jnd's are sometimes higher, sometimes lower, but the deviations of the points of subjective equality are consistently larger, viz. between 1.9 and 3.5 semitones (they varied between 1.5 and 2.3 semitones with the pitch distance discriminators).

Since these subjects seem to have applied another criterion, their real performance needs another way of analyzing their data, considering that their scores were meant as "final pitch of second stimulus higher in Ia, lower in Ib," rather than "second movement larger." Accordingly, percentages of right-column answers have been calculated as a function of the difference in final pitch. The result is shown in Figs. 6 and 7. The just noticeable difference of final pitch in the case of rises was found to be 1.4 semitones, somewhat less than the just noticeable difference of pitch distance as found in the other groups of subjects (overall); in the case of falls, a value of 2.6 semitones was obtained, 0.5 semitone more than with the pitch distance discriminators.

B. Piano tone intervals

The outcome of experiments IIa and IIb with the piano tone intervals was essentially the same as that of the speech experiments. Main differences were the following: The percentage of nondiscriminators was lower for piano falls than for speech falls. The percentage of final pitch discriminators was higher for piano rises than for speech rises, and this was opposite with falls.

Table IV gives the results for the average pitch-distance discriminator. In general, the jnd's are less

TABLE III. Just noticeable differences and points of subjective equality for the average final pitch discriminator.

Experiment Number of subjects	Ia: Speech rises 17		Ib: Speech falls 6	
	jnd	pse	jnd	pse
Overall	2.0	+0.3	2.7	+0.3
Registers equal	0.9	+0.1	1.3	+0.2
Register up	1.6	-1.9	3.3	+3.5
Register down	1.3	+2.8	2.1	-2.6

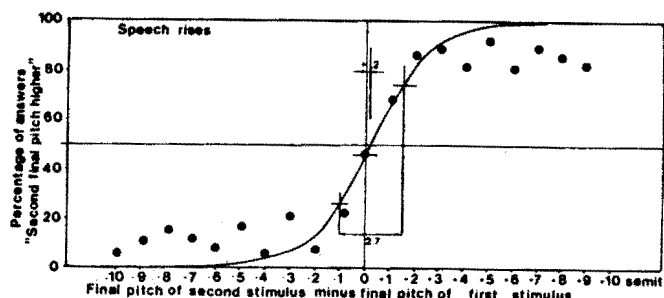


FIG. 6. Psychometric curves for the average final pitch discriminator in experiment Ia, as a function of the difference in final pitch. (17 subjects.)

than those obtained for speech. A single exception is rises with register down, jnd for speech 1.7, and for piano 1.9, which difference is no more significant than the other differences found between speech rises and piano rises. Substantial differences do occur between speech falls and piano falls in the shifting register conditions, where piano jnd's are about 1 semitone less. For both rises and falls the deviations from zero of the points of subjective equality in the shifting register conditions are smaller than those found in the speech experiments.

IV. DISCUSSION

Obviously, the task has been very difficult to many subjects. Considerable numbers of them were not able at all to perform within the limits of the experimental conditions. They had to be discarded as nondiscriminators. Among the others, the discriminators, quite a number was found to have applied a strategy that was not adequate with respect to the instruction. Falling movements appeared to be more difficult to judge than rising movements. Remarkably enough, this is not reflected in the mean jnd of the discriminators (as is suggested in an earlier report on a preliminary version of the experiments, 't Hart, 1974), but rather in the percentages of nondiscriminators, which was 16% in the experiment with rises, and 61% in the one with falls. (Corresponding figures for the piano tone experiments were 20% and 32%, respectively.)

The analysis has yielded three main effects:

(1) Rather high values for the jnd in comparison to those for the jnd of pitch in speech sounds;

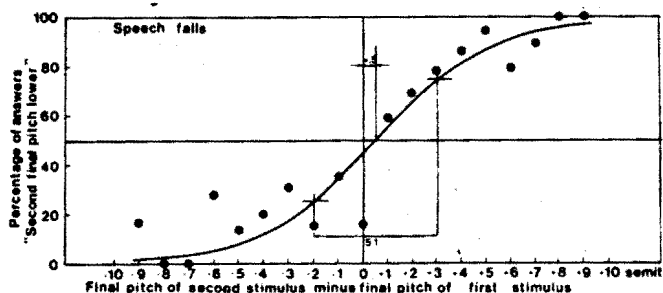


FIG. 7. Psychometric curves for the average final pitch discriminator in experiment Ib, as a function of the difference in final pitch. (Six subjects.)

TABLE IV. Just noticeable differences and points of subjective equality (pse) in semitones for the average pitch distance discriminator in the piano tone interval experiment.

Experiment Number of subjects	IIa: Piano rises 25		IIb: Piano falls 23	
	jnd	pse	jnd	pse
Overall	1.5	-0.0	1.3	-0.0
Registers equal	0.7	-0.1	1.0	+0.0
Register up	1.7	-0.9	1.6	+0.7
Register down	1.9	+1.1	1.3	-0.7

(2) considerable deviations from zero of the points of subjective equality (pse) in the shifting registers conditions;

(3) a tendency to only compare the two final pitches in a stimulus pair, instead of the distances between initial and final pitch in either member.

These effects suggest that, in any case, the description of F_0 variation in speech can be less accurate than would be expected on the basis of results of traditional psychoacoustic experiments, and yet be sufficiently precise to account for the perceptual abilities.

Meanwhile, the following remarks can be made:

(1) Best performers in the piano tone experiments had jnd's down to 0.4 semitones. This is in fair agreement with the results of Plomp *et al.* (1973) and of Attneave and Olson (1971). Upon inquiry, most good performers appeared to be musically trained.

(2) Separate concentration on the initial or the final pitch before or after a frequency glide or jump has been reported by Nabelek *et al.* (1973). Separation did most easily occur if there was a pause between the low- and the high-frequency parts of a stimulus, rather than a genuine glide. This might be applicable to the present type of stimulus material since the changes of F_0 took (partly) place in the occlusion time before a voiceless plosive, or during a voiceless fricative.

(3) The large deviations of the pse in stimulus pairs with downward-going register is of particular relevance in the case of speech, since the general downdrift of pitch in an utterance—the declination—will easily create such a relation between successive pitch movements.

Upon closer inspection of the results, it appears that it can be made plausible that the three main effects are not independent of each other.

To this end, we may try to find a possible origin of the observed deviations of the points of subjective equality. The sign of the deviations in the four different conditions is such, that if the register shift and the pitch movements have the same direction, the second member of a stimulus pair is overestimated, and if they have opposite directions, it is underestimated. The common factor can be seen in the following considerations: In stimulus pairs with objectively equally large

movements and a register shift of 3 semitones, the differences between the final pitches is also 3 semitones; in stimulus pairs corresponding to the points of subjective equality, the difference between the final pitches is always smaller. This gives rise to the supposition that at least partially also the pitch distance discriminators have based their judgments on final pitch discrimination.

This is consistent with the observation that the "genuine" final pitch discriminators show larger deviations of the pse. Furthermore, to that group were said to belong only very clear and consistent cases, and thus it is to be expected that also the group that has been characterized as pitch distance discriminators has contained several individuals who at least partially have applied the alternative strategy.

As is shown in the outcome of the consistent final pitch discriminators, the main consequence of this strategy is a deviation of the pse for stimulus pairs with shifting registers. A mixed strategy will therefore result in a curve that is obtained by taking the average of one curve with little or no deviation of the pse and a second one with large deviation. The averaged curve is less steep than each of the component curves. The jnd obtained is larger than could have been found for both "pure" pitch distance discriminators and "pure" final pitch discriminators. For instance, the curve of Fig. 4 for registers up can be obtained by taking the average of the curve for equal registers and a congruent one that is shifted 3 semitones to the left.

This fits in with the observation that final pitch discriminators did not show typically larger jnd's; it also explains why in the equal registers condition, in which both strategies are equally adequate, they are less than in the shifting registers conditions.

Overestimation of the height of the second peak in a nonsense utterance with two stressed peaks has been reported by Breckenridge (1977), more recently in Pierrrehumbert (1979). She ascribed the effect to the psychological reality of the declination (the tendency of F_0 to drift downward in pitch all over an utterance): The listener experiences the second peak which is intended by the speaker to cause an equally strong pitch accent as equally high, but due to the declination, the peak F_0 is lower. One typical result of her experiments was a difference of 8 Hz, F_0 of the first peak being 169 Hz. This corresponds to 0.8 semitones.

It seems obvious to compare this outcome with that of the present experiment Ia, for stimulus pairs with falling register. Since Breckenridge's subjects were instructed to judge which peak was higher in pitch, the most likely candidates for the comparison are the final pitch discriminators, who acted as if they had been instructed likewise.

These subjects had their pse at 2.8 semitones, which is much more than the value of 0.8 mentioned above. The 3 or 6 semitone downstep between the first and the second member of the stimulus pairs does not suggest an excessively strong declination, in view of the two

seconds time between their onsets. Thus there is no reason to suspect that this high value may originate from a steeper declination than can be found in natural speech, or in Breckenridge's stimuli.

There is another reason why the observed phenomena cannot be fully ascribed to the effect of the psychological reality of the declination alone. This effect can be restated as follows: The declination line provides a natural reference for judging pitch changes away from it; comparison of an early pitch jump with one later in the utterance is done by computing their size with reference to the declination line, irrespective of the actual pitch before the second rise. This formulation of the declination effect makes it possible, also to include the 'equal registers' and 'register upward' conditions in these considerations. It would not only predict a relatively small shift of pse in the register down condition, but also a considerable shift with registers equal, and a larger shift with upward going register. These data, however, show a negligible shift in the case of equal registers, and the shift with upward going register is smaller than that with downward going register.

Nevertheless, it does not seem necessary to abandon the declination effect entirely, in favor of an interpretation solely on the basis of "absolute peak pitch comparison" (and something comparable to it in the case of falling movements). An alternative possibility is that the value of 2.8 semitones is obtained as a result of final pitch discrimination and, in addition to that, the effect reported by Breckenridge, due to declination. In that case, one would expect substantial differences between the pse shift with stimulus pairs with upward going register and that with pairs with downward going register, since in the former condition the two effects work in opposite directions and will therefore compensate each other to a certain extent. The situation with upward going register is comparable to one in an unpublished paper by Breckenridge and Liberman (1977), in which the frequency of the valley between the two peaks was varied. When it was 2 semitones higher than the initial frequency, the declination effect was reduced to less than half. One might estimate that with a difference of 3 semitones it would have been almost annihilated. So, if the 2.8 semitones found with downward going register can be considered to be the sum of 2 (due to final pitch discrimination) plus 0.8 (due to the declination effect), one would expect 2 minus very little for upward going register. The actual outcome was 1.9 semitones.

The same trend, but less pronounced, is visible in the pse values found for the pitch distance discriminators, have only partly applied final pitch discrimination: -1.5 semitones with register up, +1.9 with register down.

Breckenridge's experiments did not comprise a condition similar to that with falling pitch movements, at least not in the sense that the pitch after the fall would be of any significance to the listener. Yet, if the declination effect has a psychological reality, it has also come into play in the experiment with falling pitch movements. This time, the final pitch discrimination effect and the declination effect will have cooperated in

the condition with upward going register and they will have counteracted in that with downward going register. So, the expectation is that the pse shift is larger in the former condition. This is confirmed in the outcome for the final pitch discriminators: +3.5 semitones with register up, -2.6 with register down. It is not confirmed in the pse values of the pitch distance discriminators, where the shift with register down is larger than with register up. This exception cannot be clarified on the basis of the assumptions made. Neither is it clear why the final pitch discrimination effect is 0.7 semitones larger with falling movements than with rises.

A final check is provided by the piano tone experiments, where the declination effect is expected not to come off. This should result in practically the same absolute values for the pse shifts, no matter the register going up or down. Taking into account the accuracy with which the pse's have been determined, we find that this is not contradicted by the results.

V. CONCLUSION

Mainly because we feel that speech pitch can best be described in terms of pitch movements rather than pitch levels, we attempted to measure jnd's of the size of pitch movements in conditions representative of the dynamic situation in speech, rather than to rely on measurements of the jnd of static pitch only. The outcome is rather complicated. In general subjects fall into three categories: One is that of the nondiscriminators, unable to discriminate differences of less than 4 semitones; a second consists of final pitch discriminators, who consistently base their judgments on the final pitch rather than assessing the magnitude of the pitch movement; the third category was called that of the pitch distance discriminators, since these subjects have based their judgments on comparisons of the pitch distances. Subjects in the latter two categories have a mean overall jnd of about 2 semitones (± 1), not only in the two speech conditions, but likewise with the piano tones.

The degree of difficulty appears to be reflected in the percentage of nondiscriminators rather than in the average performance of the discriminators. From a comparison of speech rises and speech falls in Fig. 3, we observe that subjects are better able to judge speech rises than falls. This ability could originate from the more intensive exposure to rises owing to the preference, in Dutch intonation, for the use of rises for pitch accents. If this were true, the reverse effect should occur in speakers/listeners of British English, which has a preference for falls (Meinhold, 1972).

Final pitch discriminators lack the ability of an "absolute judgment of pitch distance." In the detour necessary to accomplish the task, they come into conflict with memory capacity. For these listeners, the perceptually relevant property in the comparison of pitch movements turned out to be the pitch reached at termination of each of the movements. The accuracy with which they could discriminate the final pitches was 2

semitones on the average. This is not different from and in any case not better than the accuracy with which pitch distance discriminators did their task. Thus we may try to draw our conclusions irrespective of the strategy applied and irrespective of the question whether it is pitch distance or final pitch that is the relevant cue.

It may be clear that, under the circumstances explained above concerning the large spread and the considerable deviation of the points of equal perception, these measurements do not lead to a simple rule of thumb for judging the accuracy with which a description of speech pitch should account for the variability in the size of the movements. Nevertheless, we can say that an accuracy of 1 semitone is sufficient even in cases where final pitch matching would not give incorrect responses. But such an accuracy is hardly ever necessary, because for communication purposes the just noticeable difference is far from effective, whereas the effects of the deviation of the F_0 's apparently obscure the issue quite considerably. Thus, not until the amounts of change of F_0 in two successive pitch movements in the same direction differ by at least 3 semitones, may we expect these movements to be heard with any certainty as differently large, and even then not by every listener. This would imply that, to cover the full range from the smallest perceivable to the largest occurring change of the fundamental frequency in the intonation of Dutch, no more than four intermediate steps are required.

ACKNOWLEDGMENTS

The author wishes to thank J. de Labije and J. G. Beerends and O. H. van Hees for their respective shares in supplying subjects and in carrying out the experiments, and B. Leshowitz and B. L. Cardozo for their helpful comments on an earlier draft of the manuscript.

- Attneave, F., and Olson, R. K. (1971). "Pitch as a medium: A new approach to psychophysical scaling," *Am. J. Psychol.* **84**, 147-165.
- Bachem, A. (1937). "Various types of absolute pitch," *J. Acoust. Soc. Am.* **9**, 146-151.
- Breckenridge, J. (1977). "The declination effect," *J. Acoust. Soc. Am. Suppl. 1*, **61**, S90.
- Breckenridge, J., and Liberman, M. Y. (1977). "The declination effect in perception," (unpublished paper).

- Burns, E. M. (1974). "In search of the Shruti," *J. Acoust. Soc. Am. Suppl. 1*, **56**, S26.
- Di Cristo, A. (forthcoming). "Les traits prosodiques de l'appel en français—indices acoustiques et perceptifs," in *Proceedings of the Eighth International Congress of Phonetic Sciences*, Leeds, 1975.
- Flanagan, J. J., and Saslow, M. G. (1958). "Pitch discrimination for synthetic vowels," *J. Acoust. Soc. Am.* **30**, 435-442.
- 't Hart, J. (1974). "Discriminability of the size of pitch movements in speech," *IPO Annu. Prog. Rep.* **9**, 56-63.
- 't Hart, J., and Cohen, A. (1973). "Intonation by rule: A perceptual quest," *J. Phon.* **1**, 309-327.
- 't Hart, J., and Collier, R. (1975). "Integrating different levels of intonation analysis," *J. Phon.* **3**, 235-255.
- Isačenko, A. V., and Schädlich, H.-J. (1970). *A model of standard German intonation* (Mouton, The Hague, Paris).
- 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.
- Lehiste, I. (1970). *Suprasegmentals* (M.I.T., Cambridge).
- Meinhold, G. (1972). "Probleme der Intonationsstatistik," in *Proceedings of the International Symposium on Intonology*, Prague 1970, Acta univ. Carol. Philologica **1**, 165-173.
- Nordmark, J. O. (1968). "Mechanisms of Frequency Discrimination," *J. Acoust. Soc. Am.* **44**, 1533-1540.
- Pierrehumbert, J. (1979). "The perception of fundamental frequency declination," *J. Acoust. Soc. Am.* **66**, 363-369.
- Plomp, R., Wagenaar, W. A., and Mimpfen, A. M. (1973). "Musical interval recognition with simultaneous tones," *Acustica* **29**, 101-109.
- Pollack, I. (1952). "The information of elementary auditory displays," *J. Acoust. Soc. Am.* **24**, 745-749.
- Rakowski, A. (1971). "Pitch discrimination at the threshold of hearing," in *Proceedings of the Seventh International Congress on Acoustics*, Budapest, 1971, Vol. 3, 20H6, 373-376.
- Révész, G. (1946). *Inleiding tot de muziekpsychologie* (Noord-hollandse Uitgeversmaatschappij, Amsterdam).
- Ritsma, R. J. (1965). "Pitch discrimination and frequency discrimination," in *Proceedings of the Fifth International Congress on Acoustics*, Liège, B22.
- Rossi, M. (1971). "Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole," *Phonetica* **23**, 1-33.
- Rossi, M., and Chafcouloff, M. (1972). "Recherche sur le seuil différentiel de fréquence fondamentale dans la parole," *Travaux de l'Institut de Phonétique d'Aix*, **1**, 179-185.
- Shower, E. G., and Biddulph, R. (1931). "Differential pitch sensitivity of the ear," *J. Acoust. Soc. Am.* **3**, 275-287.
- Stevens, S. S., and Volkman, J. (1940). "The relation of pitch to frequency: a revised scale," *Am. J. Psychol.* **53**, 329-353.
- Stumpf, C. (1890). *Tonpsychologie* (Verlag S. Hirzel, Leipzig).
- Willems, L. F. (1966). "The Intonator," *IPO Annu. Prog. Rep.* **1**, 123-125.
- Witting, C. (1962). "A method of evaluating listeners' transcriptions of intonation on the basis of instrumental data," *Lang. Speech* **5**, 138-150.