Perceptual Aspects of Dutch Intonation.

J. W. M. Verhoeven

Phd
University of Edinburgh
1991

TABLE OF CONTENTS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

The Discrimination of Fall Alignment in the Dutch Hat Pattern............................154

CHAPTER 6

The Discrimination Threshold of Pitch Movement Alignment...............................172

CHAPTER 9

Declaration.

I declare that this thesis is composed by myself and that the experimental work reported is entirely my own.

Jo Verhoeven

ABSTRACT

The experimental work reported in this dissertation is aimed at investigating the perceptual relevance of pitch movement alignment to the characterization of the hat pattern, which is one of the most frequently used intonation patterns in Dutch. After a general discussion of the theoretical issues involved, the alignment of rising pitch movements in the hat pattern is examined by means of a 2AX discrimination experiment and associated labelling task. The results of these experiments provide no indication of a reliable discrimination of rise-alignment, nor do they provide evidence that the hat pattern is categorized on the basis of rise-alignment differences.

Subsequently, the discrimination of falling pitch movements in the same intonation pattern is investigated. The results of this experiment are compatible with a categorization of falling movements in terms of an early and a late category. On the basis of these findings it is concluded that the precise location of the falls is relevant to the perceptual identity of the hat pattern. The specification of rise-alignment, however, is not relevant in this perspective.

Finally, the just noticeable difference of both rise and fall-alignment is established by a discrimination experiment, which shows a differential discrimination threshold for both types of pitch movements. This threshold difference is accounted for in terms of psycho-acoustic and linguistic factors. Two follow-up studies confirms this finding, but rules out some of the psycho-acoustic hypotheses.

ACKNOWLEDGEMENTS.

Submission of a doctoral dissertation is the final stage of an institutionalized educational process and therefore constitutes the synthesis of the various educational impulses acquired during these years. Having completed this formal education, I feel that I owe a great deal to many of my teachers.

I am grateful to Prof. Emeritus *G. Jo Steenbergen* (Antwerp University) who introduced me to General and Dutch Linguistics. His blissful enthusiasm combined with exceptional pedagogical qualities have undoubtedly had a major impact on my interest in these subjects. I also express my gratitude to Prof. *René Collier* (Institute of Perception Research), who at the time was lecturing in General and Dutch phonetics at Antwerp University. These lectures were intellectually highly stimulating and truly inspirational. Their direct influence on the subject of this dissertation cannot be denied. On a more personal level, I thank him for his continuous support and friendly advice on many matters.

I also express appreciation for the teaching staff of the Department of Linguistics at Edinburgh University, who have sharpened my appetite for phonetics. I am particularly grateful to Bob Ladd, who has supervised this thesis. I admire his brilliant mind and his deep understanding of the subject elaborated in this work. I thank him for investing so much time and energy in guiding me through the Phd process.

The experimental work of this thesis has been facilitated by the help from the competent technical support staff of the Department of Linguistics. *Norman Dryden* has written various computer programmes, which have made the preparation of the stimuli for the experiments a great deal more efficient. *Cedric MacMartin* and *Stewart Smith* did an excellent job on the tape recordings.

I am also indebted to the people who have recruited the informants for the experiments and provided me with suitable laboratory facilities at their institutions: *Kris Van de Poel* at Aarhus Universitet, Afdeling Nederlandsk (Aarhus, Denmark). *Jos Wilmots* at the Economische Hogeschool Limburg (Diepenbeek, Belgium). *Daniël Bauwens* and *Mark De Kegel* at the Rijkshogeschool voor Vertalers en Tolken (Brussels, Belgium). *René Collier* at Universiteit Antwerpen, Department Germaanse Filologie (Antwerp, Belgium). *Victor Claes* at Universiteit Antwerpen, Faculteit Toegepaste Economische Wetenschappen.

Financial support for this work by a Faculty of Arts Scholarship of Edinburgh University is gratefully acknowledged.

Finally, I would like to express my sincere gratitude to my wife, Kris. She has been a source of great inspiration and it has felt very reassuring to have her full support during these eventful years.

Edinburgh, November 1990
Oxford, September 1991

CHAPTER 1

Physiological and Psycho-acoustic Aspects of Intonation.

0. Introduction.

In this chapter several fundamental issues concerning the physics and psycho-acoustics of intonation are discussed. First, a working definition of intonation is proposed, in which the linguistic concept of intonation is related to speech melody in the physical domain. Subsequently, the physiological, acoustic and perceptual dimensions of speech melody are dealt with. In a physiological perspective, speech melody is correlated with the rate of vocal fold vibration in the phonatory process. After a brief description of the anatomical structure of the larynx, the aerodynamic-myoelastic theory of vocal fold vibration is discussed. This will lead to a description of the factors determining rate of vocal fold vibration. Next, the extent that these factors contribute to the production of various aspects of speech melody are examined. Finally, a number of physiological restrictions on the production of melodic variations are pointed out. It is shown from an acoustic perspective, how speech melody is manifested in the speech signal as fundamental frequency. From a perceptual point of view, some general psycho-acoustic aspects of melodic variations in acoustically non-complex signals are discussed. Finally, the perception of pitch in more complex, speech-like signals is investigated.

1. Towards a Definition of Intonation.

In a very general perspective, intonation is an essentially linguistic concept which refers to speech melody. At the physiological level, speech melody originates from variations in the rate of vocal fold vibration during the phonatory process, which manifest themselves acoustically as variations in fundamental frequency (F0) in the speech signal. These in turn correspond auditorily to the perception of pitch. From a purely phonetic point of view, a definition of intonation along each of these dimensions can be attempted, but any such attempt poses characteristic difficulties. It is indeed fairly simple to indicate the phonetic correlates of intonation. The essential problem however lies in the fact that neither of these phonetic correlates is uniquely related to intonation.

From a physiological point of view, intonation could be defined as variation in the rate of vocal fold vibration. Vocal fold vibration is however by no means related to intonation only, since changes in the rate of vocal fold vibration also correlate with voice quality differences (modal voice vs. falsetto) or in an extreme view with the production of segmental aspects of speech such as glottal stops. In an acoustic perspective, intonation could be defined as F0 variation. Again, such a definition is too broad, since F0 is not only related to intonation. F0 variation in the acoustic signal is also related to prominence, i.e. the impression that makes some syllables of an utterance stand out with respect to other syllables. Moreover, in tone languages, F0 is related to tone rather than intonation. At the level of speech perception, similar observations can be made, in that the correlate of intonation is pitch, whereas pitch is also related to accent and tone.

Due to these intricate cross-relationships, a purely phonetic definition of intonation is impossible to formulate accurately. As a result, supplementary criteria have to be taken into consideration, such as for instance the size of the association domain or the function of intonation. These aspects are essentially phonological in nature. A working definition of intonation in this spirit can be formulated as follows:

> Intonation is the systematic use of pitch variation for communicative purposes which involves formal organization at the utterance level.

In this definition, the linguistic concept of intonation is related to pitch variation in the phonetic domain. As such, we explicitly adhere to a perceptual approach to intonation. It furthermore indicates that these pitch variations are systematic, as indicated in Pike (1972):

> The changes of pitch which occur within a sentence are not haphazard variations. The patterns of variation, the rules of change, are highly organized. Their intricacy is so great that, although one speaks his language with little effort, their analysis is extremely difficult and may induce one to conclude that no actual organization or rules are present, but that people use pitch by whim and fancy. In each language, however, the use of pitch fluctuations tends to become semi-standardized, or formalized, so that all speakers of the language use basic pitch sequences in similar ways under similar circumstances (Pike, 1972: 53).

In addition, the definition refers to the size of the phonological domain in which intonation operates. As a result, the definition distinguishes between intonation and tonal phenomena in tone languages, in that the distribution domains of intonation contours are phrases. The domain of tonal phenomena is the syllable.

Finally, the definition emphasizes the communicative function of intonation. From this functional point of view, intonation is used to signal both modality and focus effects. The former conveys speaker-related information in that deviations from the speaker's habitual pitch range may signal emotion or boredom. The latter relates to the organisation of the discourse in terms of given and new information, turn-taking etc. The investigations reported in this dissertation are however primarily concerned with the formal organisation of intonation and consequently abstraction is made from intonational function.

2. The Physiological Nature of Speech Melody.

In this section, the physiological nature of speech melody is discussed. First, the anatomical structures involved in the production of speech melody are briefly examined. Subsequently, the factors determining the rate of vocal fold vibration are reviewed. Finally, an insight is provided into how these factors interact to control speech melody.

2.1. Functional Anatomy of the Larynx.

2.1.1. Laryngeal Cartilages.

The larynx can be regarded as a complex structure of cartilages, which constitutes the expanded upper part of the trachea. It consists of nine cartilages, which are interconnected by means of the laryngeal musculature and ligaments. Only three cartilages are directly relevant in speech production: the cricoid (unpaired), the thyroid (unpaired) and the arytenoids (paired). The relationship between these cartilages is illustrated in figure 1.1 on next page:

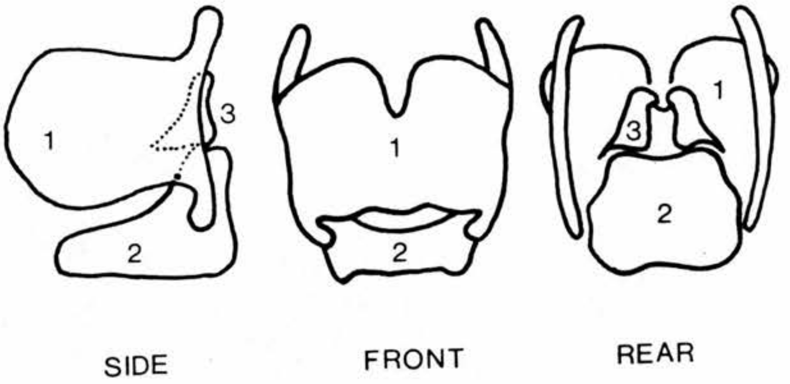


Figure 1.1: Different views of the laryngeal cartilages, which are involved in the phonatory process. 1 = thyroid cartilage; 2 = cricoid cartilage; 3 = arytenoid cartilages. (Adapted from Laver, 1980).

The inferior part of the larynx is constituted by the *cricoid* cartilage, which is shaped like a signet ring. The narrow part of this ring is situated anteriorly and it widens progressively in the posterior direction so that its upper margin extends deep into the thyroid.

The lateral walls of the larynx are formed by the *thyroid* cartilage. It consists of two quadrilateral laminae, which join anteriorly in the lower two-thirds. Each lamina ends in a vertical posterior margin, which extends superiorly and inferiorly to form the horns (cornua) of the thyroid cartilage. The superior horns are attached to the hyoid bone by the thyrohyoid ligaments, whereas the inferior horns articulate in the postero-lateral plane with the cricoid cartilage in the crico-thyroid joint. This joint enables the posterior part of the cricoid and the attached arytenoid cartilages to swing towards or away from the anterior part of the thyroid cartilage, thus increasing or decreasing the horizontal dimension of the larynx. The thyroid cartilage itself seldom performs a rotating movement along this joint, since it is firmly anchored to the surrounding skeletal parts (Sonessen, 1974).

Finally, the arytenoid cartilages are pyramid shaped. Each extends laterally in the muscular process and projects anteriorly to form the vocal process, to which the vocal folds are attached. The arytenoids articulate with the cricoid by means of the crico-arytenoid joints, which allow rocking and gliding movements of the arytenoids. The rocking movement involves a rotation of the arytenoids round their vertical axes, which causes their vocal processes to swing laterally or medially. This movement abducts or adducts the vocal folds. In the gliding movement, the arytenoids slide transversely on the lamina of the cricoid cartilage, enabling them to move closer together or further apart. The effect of both movements is illustrated in figure 1.2:

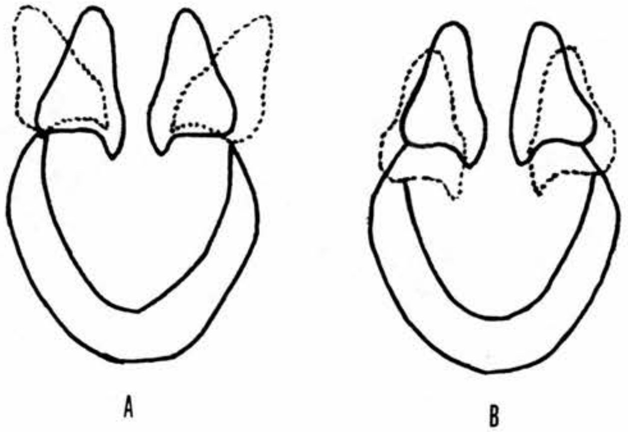


Figure 1.2: Illustration of the rocking (a) and gliding (b) movements of the arytenoid cartilages (Adapted from Clark & Yallop, 1990).

2.1.2. Laryngeal Cavity.

The cavity of the larynx is divided into a superior and inferior part by the vocal folds, which anteriorly originate from the thyroid lamina, below the thyroid notch, and insert posteriorly on the vocal process of the arytenoid cartilages. The vocal folds are essentially wedge-shaped, their apexes projecting medially into the larynx. It is important to point out here that they essentially consist of a layered structure (Hirano & Kakita, 1985). From a histological point of view, five layers can be distinguished. The wet mucous membrane of the vocal folds is subdivided into the epithelium and the lamina propria. The latter consists of three layers, depending on the density of the fibres involved. In the superficial layer, the fibrous components are very loose and resemble a mass of soft gelatine. In the intermediate layer, the fibres are elastic, resembling soft rubber bands. The deep layer consists of collagenous fibres, resembling cotton thread. Consequently, there is a gradual increase in stiffness from the superficial to the deep layer. Finally, there is the main body of the vocal folds, i.e. the vocalis muscle, which looks like a set of rubber bands. From a mechanical point of view, Hirano and Kakita (1985) classify these layers into three categories. First, there is the *cover*, which consists of the epithelium and the superficial layer of the lamina propria. The intermediate and the deep layers of the lamina propria constitute a *transition* to the vocalis muscle, which constitutes the *body*. The mechanical properties of this layered structure have important consequences for the pattern of vibration of the vocal folds, which will be described in section 2.2.

Above each vocal fold, there is a ventricular fold, which is separated from the corresponding vocal fold by a small horizontal groove: the ventricle of Morgangi. These ventricular folds consist of mucous glands and a few muscle fibres. They are situated wider apart than the vocal folds and play no role in phonation.

An anterior section of the laryngeal cavity is given in figure 1.3:

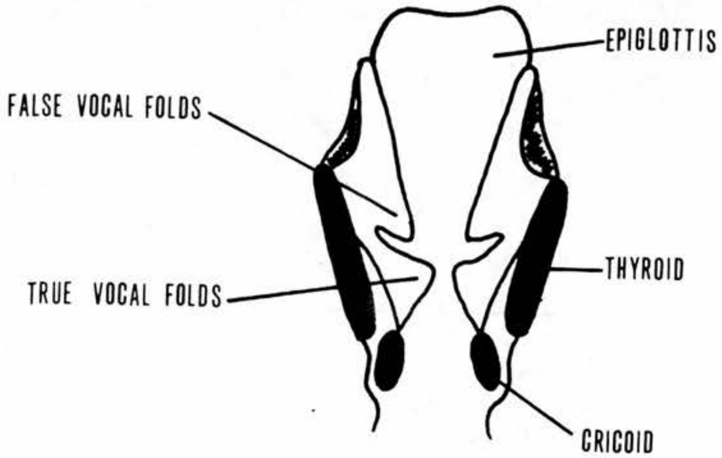


Figure 1.3: Anterior view of the laryngeal cavity (Adapted from Clark & Yallop, 1990).

2.1.3. Laryngeal Musculature.

The muscles of the larynx are traditionally subdivided into extrinsic and intrinsic laryngeal muscles, depending on whether both insertion points are situated on the larynx (intrinsic) or on whether one point of insertion is located outside the larynx (extrinsic). From a functional point of view, the action of the intrinsic muscles determine the mechanical operating characteristics of the larynx: by changing the position of the laryngeal cartilages relative to each other the intrinsic musculature determines the length of the vocal folds, their medial compression and their position. The extrinsic musculature is submitted to a further functional subdivision by Hardcastle (1976) into laryngeal elevators, the action of which raise the larynx, and laryngeal depressors, which lower the larynx. The following will confine itself to a brief description of the intrinsic musculature. The information given here relies on anatomical descriptions in Sonesson (1974), Romanes (1978), Daniloff et al. (1980), Laver (1980) and Hirano & Kakita (1985). Although all the intrinsic laryngeal muscles are paired, they are referred to here in the singular form.

The *cricothyroid* muscle (*pars recta* and *pars obliqua*) originates from the anterior lateral outer surface of the cricoid. The pars recta inserts on the inferior rim of the thyroid lamina, whereas the pars obliqua attaches to the inferior horns of the thyroid. Contraction of the pars recta lifts the anterior part of the cricoid cartilage upwards and as a result, its posterior portion tilts downwards. Contraction of the pars obliqua has a similar effect, but also tends to slide the thyroid slightly forwards. Consequently, the action of the cricothyroid increases the distance between the insertion points of the vocal folds and consequently elongates them. This makes the vocal folds thinner and sharpens their edges.

The muscle involved in glottal relaxation is the *thyro-arytenoid*, which arises from the posterior surface of the thyroid cartilage close to its midline. It inserts on the anterolateral surface of the arytenoid cartilage and constitutes the main body of the vocal folds. Contraction of this muscle pulls the arytenoids forward, which shortens and thickens the vocal folds. Furthermore, their edges are rounded.

The *posterior crico-arytenoid* muscle is the major abductor of the vocal folds. It originates from the posterior surface of the cricoid lamina and attaches to the muscular process of the arytenoids. Contraction of this muscle rotates the arytenoids round their vertical axis, thus swinging their vocal processes laterally. It elongates and thins the vocal folds and rounds their edges.

The major adductor muscle is the *lateral crico-arytenoid*, which originates from the lateral anterior part of the cricoid arch and inserts on the muscular process of the arytenoid cartilages. Action of this muscle pulls the muscular process of the arytenoids anteriorly and rotates the arytenoids thus closing the glottis. It also lengthens and thins the vocal folds to some extent.

Finally, the *interarytenoid* muscle consists of two functional parts: the *transverse* and the *oblique*. The fibres of the transverse arise from one arytenoid cartilage and course horizontally to insert on the posterior surface of the other arytenoid. Contraction causes the arytenoids to slide towards each other. The oblique interarytenoid has a characteristic X-shape. Each belly courses from the muscular process of one arytenoid towards the apex of the other. Contraction of this muscle also slides the arytenoids towards each other and simultaneously tips them a little. The interarytenoid slightly shortens, thickens and slackens the vocal folds.

The functional effects of these muscles are illustrated in figure 1.4:

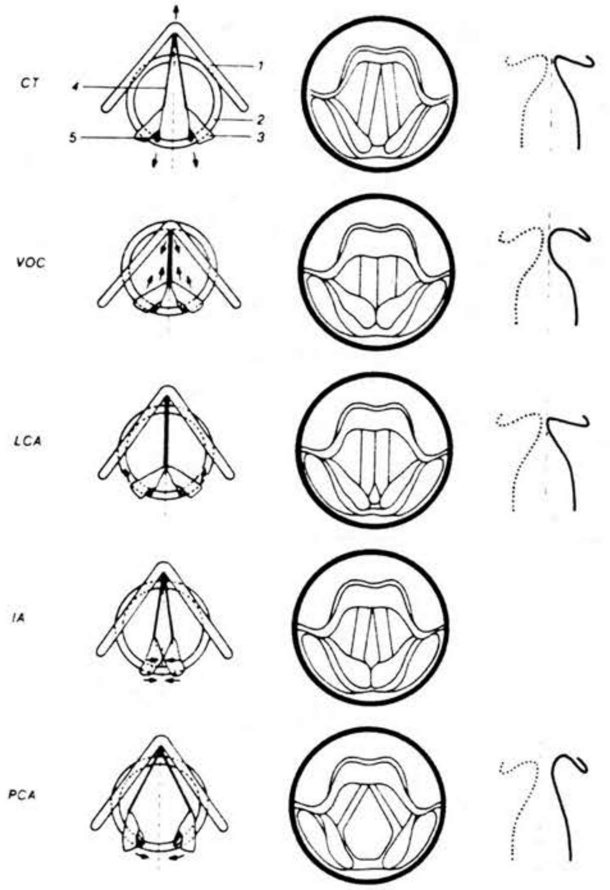


Figure 1.4: Diagrammatic representation of the effects of muscle contraction on the position and shape of the vocal folds. Arrows in the left-hand column indicate the direction of the muscular force. 1=Thyroid, 2=Cricoid, 3=Arytenoid, 4=Vocal ligament, 5=Posterior cricoarytenoid ligament. The middle column illustrates the view from above, whereas the right column presents contours of frontal sections of the vocal fold; the dotted line shows the vocal fold contour when no muscle is activated (Adapted from Hirano & Kakita, 1985).

2.2. The Aerodynamic-Myoelastic Theory.

Of the competing theories of vocal fold vibration, the aerodynamic-myoelastic theory is undoubtedly the most widely accepted account of the forces that play a role in phonation. A summary of these forces is given in Van den Berg (1974):

> (...) the function of the larynx is based on the interplay of three factors: (1) the aerodynamic properties of the air which actuates the larynx, (2) the adjustment of the larynx, brought about by the proper nervous activation of the various muscles, and the myoelastic properties of the laryngeal components, (3) the aerodynamic coupling between (a) the subglottal system and the larynx, (b) the left and right vocal fold, and (c) the larynx and the supraglottal system (pp. 291-292).

The *myoelastic* component, first postulated by Müller (1837), refers to the *adduction* of the vocal folds to interrupt the pulmonic egressive airstream and their *elasticity*, which enables them to be stretched or relaxed in the production of pitch changes. The *aerodynamic* component refers to the airflow through the glottis, which creates suitable physical conditions for the glottis to be closed. These forces are examined in greater detail with reference to a typical cycle of vocal fold vibration.

In the first stage of a vibratory cycle, the vocal folds are adducted by activation of the appropriate laryngeal muscles. This closure of the glottis interrupts the pulmonic egressive airstream and as a result, subglottal pressure starts building up to a point where it reaches a sufficiently high level to separate the vocal folds. This separation is achieved with a vertical phase difference (Laver, 1980), i.e. the lower parts of the vocal folds separate first, followed by a separation of the upper parts. Upon their complete separation, the airstream is forced through this narrow constriction, which causes the jet of air to accelerate. At this point the Bernouilli principle comes into operation, the effect of which is described in Borden & Harris (1980) as follows:

> Simply stated, the Bernouilli principle is that such an increase in velocity results in a drop in pressure exerted by the molecules of moving gas or liquid, the pressure drop being perpendicular to the direction of the flow (p. 83).

The Bernouilli effect thus causes a pressure drop in the glottis and consequently a sucking force is exerted on the lateral surfaces of the vocal folds, which is perpendicular to the direction of the airstream. As the air escapes, the subglottal pressure decreases until it is sufficiently low to be overcome by the myoelastic tension of the vocal folds and the Bernouilli effect. At this point, the glottis closes naturally and the cycle starts repeating itself.

It deserves mentioning that the fine detail of vocal fold vibration is in fact far more complex than described here, in that during vibration, a lateral wave motion in the mucosal layer can be observed as well. This wave motion propagates from the medial edges of the vocal folds towards their lateral boundaries. This wave motion is accounted for by the cover-body theory (Hirano & Kakita, 1985), which indicates that the vocal folds constitute a layered structure in which the loose mucous cover is supported by progressively stiffer deeper layers. Laver (1980) suggests that:

> During phonation, the cross-sectional shape of each vocal fold is subject to continuously changing deformation. Part of the dynamic deformation is attributable to the mucosal wave motion travelling up the external surfaces of the vocal folds and into the ventricles, and part to the more gross displacements involved in the vertical phase difference mentioned earlier (p. 99).

2.3. Determinants of Vocal Fold Vibration Rate.

The rate of vocal fold vibration in the phonatory process is determined by a number of factors, which are directly relevant to the production of speech melody. In the following, a distinction is made between primary and secondary factors.

The primary determinants of rate of vocal fold vibration are their *tension* and the *subglottal pressure* that activates them during phonation. The tension of the vocal folds is to a large extent determined by their length: as the vocal folds are stretched their tension increases, thus causing higher frequencies of vibration. This type of tension which arises from length variations is essentially passive, since it is caused by laryngeal muscle activity which changes the position of the laryngeal cartilages to which the vocal folds are attached. The relationship between the production of speech melody and length variations of the vocal folds was investigated by Hollien (1960a, b) and Hollien & Moore (1960) in a series of studies in which vocal fold length was measured under various experimental conditions.

In Hollien (1960a), the relationship between intrinsic larynx size and overall voice level was studied. For this purpose, four groups of subjects with differently pitched voices were selected. One group consisted of very low-pitched male voices, the second of very high-pitched male voices, the third of low-pitched female and the fourth of high-pitched female voices. In the experimental set-up, lateral X-ray photographs were made of their larynxes. These X-rays were carefully calibrated and four indices of laryngeal size were calculated. Thus, two anteroposterior measurements, one vertical and one area measurement were obtained. From the results, it emerges that significant differences in larynx size can be observed between the male and female group. Similar differences can be seen between the low- and high-pitched voices within each sex group. Hollien concludes that the overall larynx size, which naturally determines the intrinsic length of the vocal folds, is significantly correlated with the individual pitch level of the subjects, both within and across sexes. As such, the intrinsic length of the vocal folds determines an individual's overall pitch of voice in speech.

In Hollien (1960b), vocal fold length of the same subjects was measured during the production of pitch variations. On this occasion, subjects were required to phonate at four different pitch levels, three of which were distributed across their normal pitch register and one was situated in the falsetto register. The pitch levels were specified for each subject individually as a proportion (10, 25, 50 and 80 %) of his/her overall pitch range. This proportion was defined with respect to the subject's lowest sustainable tone. The experimental control over the pitch levels to be realized was achieved by means of a set of reference tones. During the realization of the pitch levels, the vocal folds were photographed. Subsequently, the photographs were calibrated and vocal fold length was measured.

The results of this experiment corroborate the findings of the previous one, in that highly consistent differences were found between the experimental groups. At all investigated pitches, the length of the vocal folds decreases from low to high male to low and high female voices. Furthermore, a systematic elongation was observed as the pitch increased. This elongation was not confined to specific areas of the subjects' pitch range and Hollien points out that this is unlike what had been suggested in previous literature.

Finally, it should be pointed out that Hollien also made measurements of the vocal folds in an abducted state. If these results are compared with those in phonation, it turns out that the vocal folds are without exception longest in a state of abduction. Thus, it seems to be the case that the vocal folds are always shortened during phonation in comparison to their abducted length.

In Hollien & Moore (1960), a more detailed investigation of vocal fold length was reported. A new group of male speakers with differently pitched voices were required to produce a larger variety of pitch levels. The measurements of the vocal folds in these circumstances entirely corroborate the results in Hollien (1960b).

From a physiological point of view, vocal fold length can be adjusted during phonation by changing the position of the cricoid relative to the thyroid cartilage. Tilting the cricoid upwards lengthens and stretches the vocal folds, whereas tilting the cricoid downwards shortens and relaxes them. As indicated above, greater stretching of the vocal folds is correlated with higher frequencies of vibration, whereas their relaxation causes low frequencies of vibration. This tilting action of the cricoid is achieved mainly by the crico-thyroid muscle (pars recta). Its active role in the production of speech melody was shown by Collier (1975b) and will be discussed at length in section 2.4 of this chapter. It has furthermore been observed that elongation of the vocal folds is often accompanied by activity of the pars obliqua of the crico-thyroid muscle, the result of which is a small forward shift of the thyroid relative to the cricoid.

The second conditioning factor of rate of vocal fold vibration is subglottal air pressure (henceforth P_{sg}): as P_{sg} increases, the speed with which the airstream is forced through the glottis increases and so does the force of the Bernoulli effect. As a result, the vocal folds are sucked together more rapidly, which results in a shorter closing phase. The opening phase is also shorter at higher P_{sg} levels, since the vocal folds are forced apart more quickly. The combined effect is that the vocal folds vibrate at higher rates, resulting in a higher speech melody. This effect has been shown by Ladefoged (1967) and others.

Vocal fold length/tension and subglottal pressure can be considered as the primary factors in determining the rate of vocal fold vibration and consequently the frequency of speech melody. The boundaries within which these factors influence rate of vibration are naturally determined by the intrinsic length and cross section of the vocal folds.

Besides these, a number of secondary factors related to interactions of the larynx with other systems, should be mentioned. First there is the interaction with the *subglottal respiratory system*. In normal circumstances, subglottal air pressure does not vary a great deal during each vibratory cycle. At certain frequencies of vibration however, so-called subglottal resonance occurs. At this frequency, subglottal pressure is reported to vary markedly during the glottal cycles. Lieberman (1968) suggests that:

> it is probably difficult to produce a normal glottal output when the subglottal pressure varies during each vibratory cycle (p. 33).

As a result, speakers systematically avoid phonation at a frequency that excites the subglottal resonance at 300 Hz, which is confirmed by particularly informative results of F0 measurements for a large number of speakers in Lieberman (1968). Secondly, there are interactions with the ***supraglottal*** system. In open articulations, the pharyngeal air pressure (Po) is equal to atmospheric. Po however rises quickly when the vocal tract is closed, with the result that at some point Po will equal P_{sg}, causing the airflow to stop. In less extreme cases of momentary closure or partial obstruction the difference between pharyngeal and subglottal pressure diminishes, causing the velocity of the transglottal airflow to slow down. This directly affects the force of the Bernoulli effect, which decreases proportionally. Variations of this kind typically cause sudden fluctuations or pitch perturbations associated with segmental effects, such as the production of voiced stops and onset/offset of phonation (Lehiste, 1970). Such pitch perturbations are generally regarded to be without linguistic significance, but they do seem to play an important role in the naturalness of speech melody.

2.4. Controlled Variables in the Production of Speech Melody.

The basic control mechanisms of the rate of vocal fold vibration in speech were investigated systematically by measuring muscle activity and respiratory gestures during the production of speech melody variations (Collier, 1975b). The experimental set-up was as follows. A native speaker of Dutch read a list of utterances with a large variety of acceptable pitch contours. During his deliveries, muscle activity was measured by means of traditional electro-myographic (EMG)(1) techniques. The muscles chosen for this experiment were the cricothyroid (CT), the sternohyoid (SH), the sternothyroid (ST) and the thyrohyoid (TH). In addition, simultaneous recordings of subglottal pressure (P_{sg}) were made. P_{sg} was measured directly by means of a pressure transducer which was coupled to a plastic tube that was inserted into the subglottal space of the larynx through the cricothyroid membrane. Furthermore, F0 traces were obtained for each of the subject's productions.

The main conclusion of this experiment was that the cricothyroid muscle bears the most direct relationship to F0 changes. Activity of this muscle was found to corre-

---

(1) Electromyography (EMG) is a technique in which fine wire electrodes are inserted into the muscles to record the electrical potential that is produced by muscle contraction.

late with the direction of F0 change, in that increased activity was observed in the production of rising F0 movements, while falls were accompanied by CT relaxation. Moreover, Collier observes a relationship between CT activity and the size of pitch movements (activity was stronger in large movements) and the rate of F0 change (activity mirrored the slope of pitch movements). CT was inactive only when F0 was low and gradually falling, i.e. during stretches of declination.

Regarding the other muscles under investigation, it was observed that the thyrohyoid bears no systematic relationship to F0 variations. The sternohyoid is primarily associated with segmental articulations involving tongue retraction or lowering. In certain instances however, the SH is observed to assist in the production of F0 falls, especially those occurring in the speaker's middle to low register.

As far as subglottal pressure is concerned, Collier does not find a relationship with local F0 variations which originate from CT activity. It is only the low declination line which is found to be entirely a function of gradually falling P_{sg}. This observation has given rise to the view that declination has to be regarded as a by-product of the respiratory system: as the utterance proceeds, P_{sg} falls, resulting in a progressively slower rate of vocal fold vibration which consequently causes F0 to fall gradually.

Collier's observations are corroborated by Maeda (1979), who carried out EMG measurements on selected laryngeal muscles during the production of utterances by a native speaker of American English. He found a similar relationship between F0 movements and activity of the cricothyroid as Collier did. Furthermore, the sternohyoid and sternothyroid were observed to participate in F0 lowering gestures. Unlike Collier, Maeda postulates that F0 declination is the consequence of the geometrical relationship between the cricoid and thyroid cartilages which, he suggests, is not only determined by cricothyroid activity, but also by tracheal pull. This is defined as a downward force acting through the trachea upon the cricoid as a result of decreasing lung volume. Maeda hypothesizes that as lung volume decreases, tracheal pull increases, causing a gradual tilt of the cricoid cartilage. This results in a gradual shortening of the vocal folds, which gives rise to lower frequencies of vibration. Maeda does not present any concrete evidence to support this hypothesis, but argues that only a very small downward movement of the cricoid is needed to account for the observed shortening of the vocal folds in sentence production, that is 1.3 mm.

Both Collier and Maeda share the view that sentence declination is an automatic by-product of the respiratory system, and as such is beyond the speaker's active control. This view of declination as an essentially passive phenomenon has been criticized since, and the issue was subject to further investigation in Collier (1987) and Gelfer et al. (1987). Collier (1987) investigates the relationship between the physiological control mechanisms of speech melody and two aspects of declination: its slope and its setting. The control of declination slope was investigated by having a native speaker of Dutch mimic utterances of varying length with reiterant speech using /fa/ and /ma/ syllables. The intonation contours were such as to yield three stress conditions: some utterances had an early stress, some a late stress, whereas a third group contained two stresses. In all three conditions, the stress was realized prosodically by means of an F0 rise, immediately followed by an F0 fall. During the productions, simultaneous recordings of CT activity, subglottal pressure and F0 were made, in a fashion similar to Collier (1975b).

The results of this experiment show that the slope of the declination line becomes less steep with increasing utterance length. At the same time, it is observed that subglottal pressure exhibits a negative rate of change which covaries with the declination slope. In the late stress condition, significant CT activity was measured at the onset of utterances, followed by a gradual relaxation of this muscle. When the subject was asked to produce a sequence of /ma/ syllables at a constant pitch, the level of P_{sg} remained constant. From these observations, Collier concludes that F0 declination slope can be accounted for in terms of variations in subglottal pressure. In certain instances however, the role of a laryngeal component cannot be denied, since a gradual relaxation of the CT is observed.

In the second experiment, attention was focused on the relationship between P_{sg}, laryngeal muscle activity and declination setting(2). A subject was asked to produce utterances of the type mainclause+subclause and subclause+mainclause. These utterances were designed in such a way that the same clause could occur before or after a syntactic boundary. This is illustrated in (1) and (2):

(1) Omdat hij ziek is, wil Jan in bed blijven.
(Because he is ill, Jan wants to stay in bed.)

(2) Wil Jan in bed blijven, omdat hij ziek is?

---

(2) The term declination 'setting' covers both setting of the initial F0 value of the declination line at the beginning of utterances and the resetting of the F0 value after a major syntactic boundary.

(Does Jan want to stay in bed because he is ill?)

Besides these short clauses, there were also utterances with long clauses containing roughly twice as many syllables. An example is given in (3):

> (3) Omdat hij zo vreselijk verkouden is, wil Jan maar liever in bed blijven. (Because he has such a bad cold, Jan would rather stay in bed.)

From a prosodic point of view, the speaker realized each clause with four possible pitch contours. As in the former experiments, CT activity, subglottal pressure and F0 were measured simultaneously.

At the acoustic level of analysis, the results of the experiment clearly show that the declination line is reset after a major syntactic boundary. This is to say that the speaker starts declination at an F0 level which is perceptually higher than that at the end of the first clause by raising F0. As far as the relationship between P_{sg} and declination resetting is concerned, it was found that the subject indeed raises P_{sg} slightly at a reset, but not enough to account for the reset in terms of P_{sg} alone, from which Collier concludes that the speaker only raises P_{sg} to a level which allows for adequate phonation. The onset of the declination line depends on the level of CT activity.

In the discussion so far, it has not been considered explicitly whether P_{sg} itself is a controlled variable in the production of speech melody. Two possibilities suggest themselves: the drop in subglottal pressure associated with declination can either be the passive consequence of lung deflation or alternatively there is the possibility that the time course of falling P_{sg} is carefully controlled. Both alternatives were investigated in an experiment reported in Gelfer et al. (1987), in which two native speakers of Dutch produced reiterant forms of Dutch utterances with /fa/ and /ma/ syllables, which essentially involve different airflow requirements. In the /fa/ utterances, there is a reduced airflow resistance at the glottis due to the voiceless fricatives, resulting in higher airflow rates than in utterances containing /ma/ syllables. If the lungs deflate passively, it is to be expected that subglottal pressure declines at different rates in both types of utterances. If subglottal pressure is actively and dynamically controlled, P_{sg} and airflow are expected to be unrelated.

During the production of both sets of utterances in the experiment, airflow rate was calculated as a function of lung volume variations over time. In this, lung volume was derived by means of a plethysmograph(3) . Subglottal pressure was measured by insertion of a pressure transducer in the subglottal cavity. In addition, CT activity was recorded and F0 was derived from an accelerometer attached to the pretracheal skin.

The results of these measurements are summarized in figure 1.5:

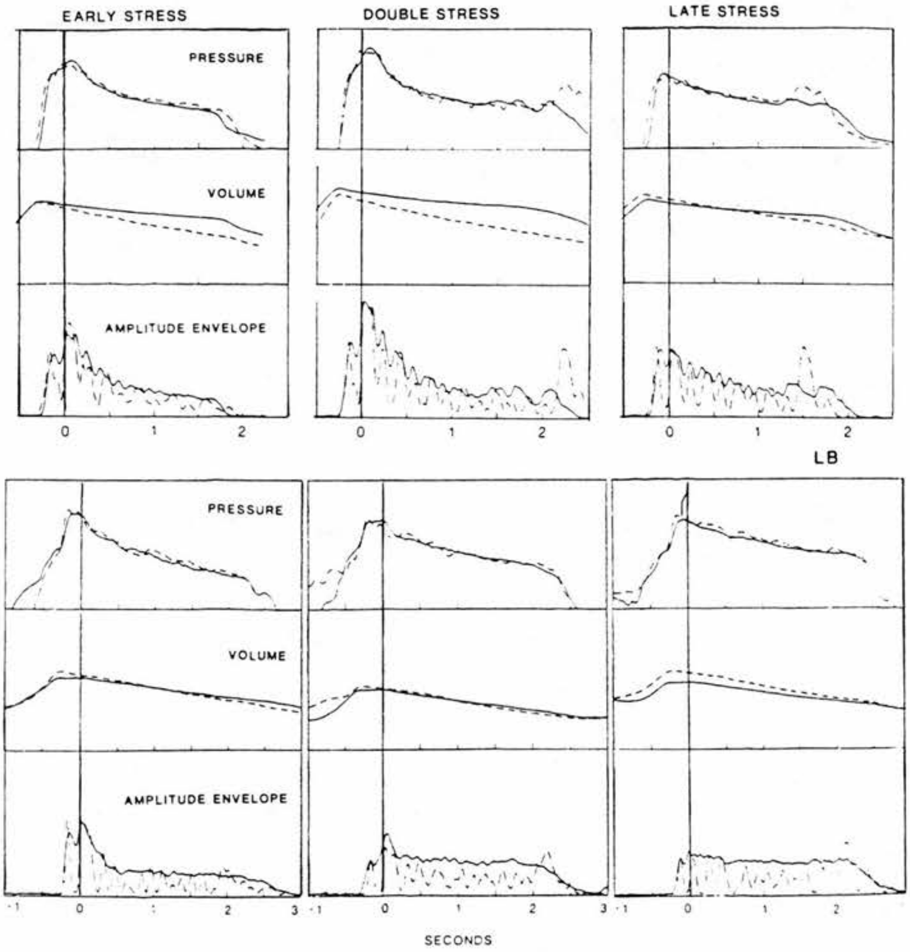


Figure 1.5: Averaged subglottal pressure, volume and amplitude envelope curves for comparable /ma/ and /fa/ utterances for two subjects RC and LB. The vertical line in each box denotes the line-up point used for averaging the tokens of each utterance type, which in these utterances is the onset of the vowel for the first syllable receiving lexical stress. The solid curves represent the reiterant /ma/ utterances, and the dashed curves the reiterant /fa/ utterances (Adapted from Gelfer et al., 1987).

---

(3) Plethysmography involves placing the subject in a tank, with his head protruding through an opening which is hermetically sealed at the neck, somewhat after the manner of an 'iron lung'. As the subject's chest expands and contracts, air is driven out of and into the tank. The volume of air thus displaced or the volume-velocity of the air-flow out of or into the tank can be measured (Catford, 1977).

It can be seen quite clearly that there are only marginal differences in the course of P_{sg} in both types of utterances. Airflow on the other hand is clearly different in that the slope of the airflow curve is steeper in /fa/ utterances. As such, it is shown that the decline in P_{sg} is essentially independent of rate of airflow, which indicates that subglottal pressure is a carefully controlled parameter in the production of speech melody.

On the basis of these experiments, Collier (1987) proposes a production model of speech melody variations where the rises and falls, which are superimposed on a gradually declining baseline, are attributed solely to activity of the cricothyroid muscle, whereas the baseline itself is mainly correlated with falling P_{sg}. A simple model like this can, according to Collier, account for a fairly large number of intonation contours in short utterances.

In longer utterances, the model has to be slightly more complex, since such sentences tend to start with higher pitch levels. This is generally achieved by CT contraction appropriate to reach the intended pitch level. The initial stretch of the declination line is then the combined effect of CT relaxation and falling P_{sg}. Upon termination of CT activity, declination can be accounted for by falling P_{sg} only.

The final point that is addressed in this section concerns the physiological restrictions on the production of speech melody. We will confine ourselves to the speed with which melodic changes can be produced, since other aspects are -to my knowledge- not investigated.

A study by Sundberg (1979) is particularly informative in connection with maximal speed of melodic change. His investigation was carried out on trained and untrained speakers of both sexes. The major conclusion from this survey is that the speed with which melodic changes can be achieved is significantly correlated with several factors, such as the direction of the melodic movement, the tonal interval, sex of the speaker and degree of training. It is also observed that melodic elevations take longer to complete than melodic falls. In untrained subjects for instance, a 4 semitone rise takes about 80 msec to complete; a similar fall about 67 msec(4). Furthermore, the response time increases slightly as the tonal gap becomes bigger. This trend is only significant in the production of rises, in which a distance of 4 semitones takes 80 msec to bridge. For a 12 semitones gap 100 msec is needed.

---

(4) The figures given here are approximate, since they had to be derived from a graphical representation of the results in the original paper.

Sex of the speaker is another important factor, in the sense that female speakers need considerably less time than males to complete both rises and falls (75 msec and 60 msec compared with 85 msec and 75 msec respectively for males). As to the factor of voice training, it was observed that the trained subjects (singers) always need less time to complete a melodic movement than untrained subjects.

From these observations, it has to be concluded that speakers need a minimum amount of time to produce melodic variations, depending on each of the above-mentioned factors. It is thus clear that it is physiologically impossible to produce sudden pitch jumps, due to constraints on the phonatory system.

3. The Acoustic Nature of Speech Melody.

The acoustic correlate of speech melody is the fundamental frequency (F0) of the speech signal. F0 can be regarded as a physical characteristic of the speech waveform. The source-filter model of speech production (Fant, 1960) indicates that any speech output can be described in terms of the combined action of a source and an acoustic filter. In the production of normal speech, two types of sources can be distinguished. In voiced speech, the sounds are generated at the glottis by vocal fold vibration. In voiceless speech, the sounds are generated elsewhere by bringing two articulators in a state of close approximation, so that the airstream from the lungs is forced through a very narrow gap, causing a turbulent airstream. In both instances, the ultimate quality and identity of the speech output is determined by the shape of the acoustic filter, which is constituted by that part of the oral cavity between the source and the outside air. In voiced speech, the filter consists of the pharynx and the oral cavity (in nasal sounds the nasal cavity is involved as well). In voiceless speech, the filter consists of the oral cavity between the location of the articulators in close approximation and the lips (inclusive). In the following, we confine ourselves to a brief discussion of voiced speech sounds.

In voiced speech, the filter is driven by acoustic energy produced by vocal fold vibration. Each time the vocal folds open and close, a puff of acoustic energy is sent into the filter, and causes a momentary rise in pressure with respect to atmospheric pressure. These pressure variations give rise to sound and can be represented as a waveform. There are several types of waveforms, a brief digression into which is appropriate here. The simplest of waveforms is a sine wave, an example of which is given in figure 1.6:

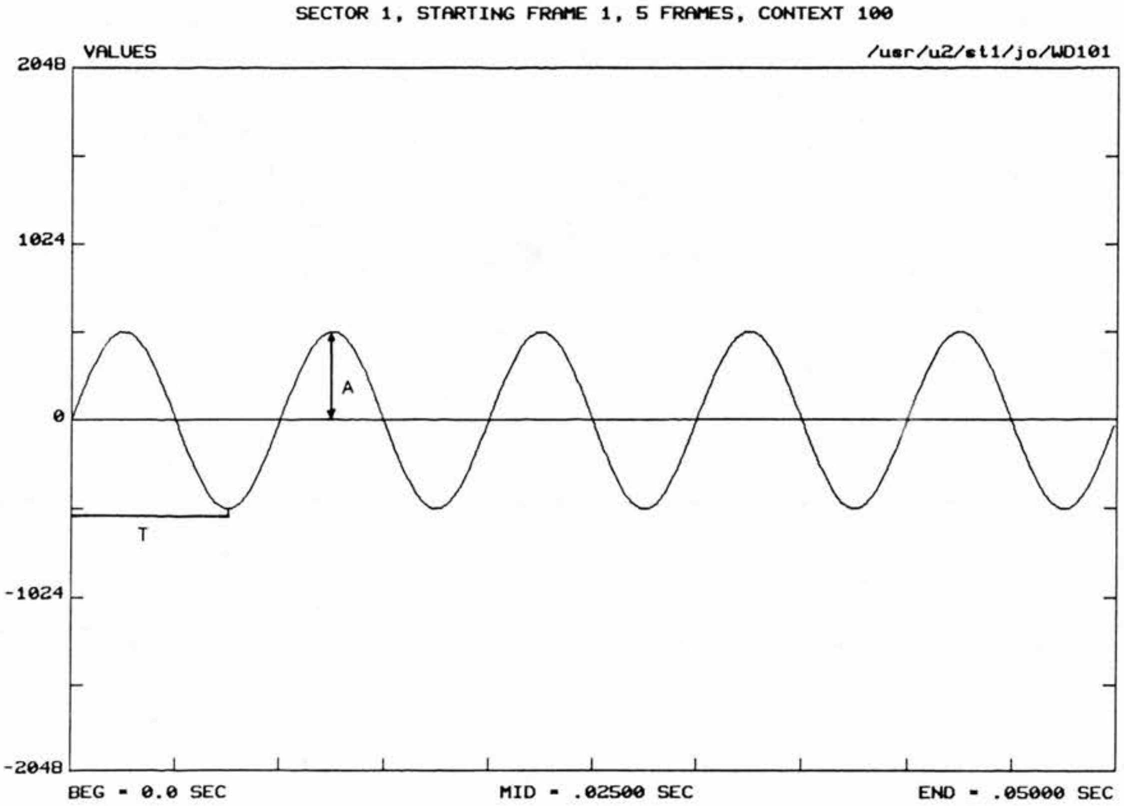


Figure 1.6: Illustration of the physical characteristics of a sine wave. A = amplitude and T = period. There are five cycles in 0.05 sec so that the frequency is 100 cycles per second, i.e. 100 Hz.

Such a waveform can be physically characterized in terms of its amplitude, its period and its frequency. The amplitude (A) of the waveform can be defined as its maximal displacement from the base-line. If the wave is the representation of sound, amplitude is measured in decibel (db) and correlates with sound intensity (loudness). The period (T) of a waveform is the time that elapses before the wave repeats itself. The frequency (F) of the waveform refers to the number of times the wave repeats itself in a given time. In physiological terms, it correlates with the number of times the glottis opens and closes per second. Frequency is measured in Hertz (Hz) which is an expression of the number of times the wave repeats itself per second. Frequency and period of a sine wave maintain a fixed relationship, which can be expressed in the formula T=1/F. In the above example, the wave repeats itself 100 times per second, and, consequently, its frequency is 100 Hz. Its period can be derived from this by T=1/100 = 0,01. It can be concluded therefore that each sine wave is uniquely defined in terms of its amplitude and its frequency. For completeness sake, it should be mentioned that two otherwise similar sine waves can also differ in their phase, i.e. the time displacement between the two waveforms. Since the human ear is not very sensitive to phase differences (Nooteboom & Cohen, 1984), these will be left out of consideration.

All waveforms which are not sinusoidal in shape are complex waveforms. Here, a distinction has to be made between periodic and aperiodic waveforms. The former repeat themselves in time, whereas the latter do not. It is a fundamental characteristic of periodic complex waveforms that they can be mathematically represented as the sum of a number of sine waves. These sine waves represent the frequency components of the corresponding sound. The result of a mathematical decomposition of a complex wave into its constituting components is given in figure 1.7:

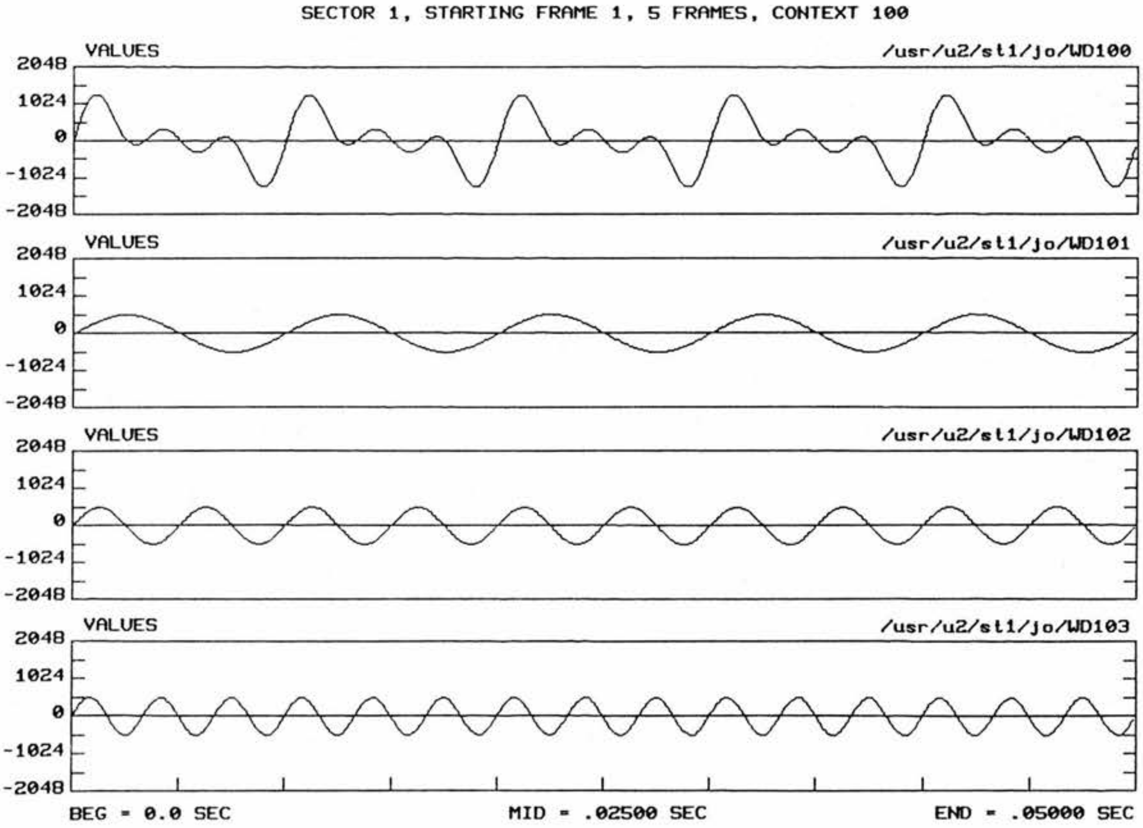


Figure 1.7: Decomposition of a complex periodic waveform with a frequency of 100 Hz (top) in its constituting frequency components of 100, 200 and 300 Hz. All components have the same amplitude.

It can be observed that the frequency of the complex waveform in figure 1.7 is equal to the frequency of its lowest component. This component is called the fundamental, whereas the other components are known as the harmonics. A further characteristic of periodic complex waves is that the fundamental and the harmonics maintain a specific relationship, in the sense that the frequencies of all the harmonics are whole number multiples of the frequency of the fundamental. The result of this is that all complex periodic waveforms can be decomposed into their constituting frequency components by a mathematical technique known as Fourier Analysis (after the French mathematician Fourier). This analysis provides us with an insight in the frequency components that are typical for speech sounds, the results of which are traditionally given in the form of a spectrum. In the spectrum, the frequency of the fundamental represents the source characteristics in the source-filter model. These source characteristics are responsible for F0 in the speech signal, which is a direct reflection of rate of vocal fold vibration. The harmonics represent the filter characteristics, i.e. their frequencies are responsible for the quality of speech sounds.

The detection and estimation of fundamental frequency in speech signals has received considerable attention. Hess (1982) makes a taxonomical distinction between F0 detection algorithms which work in the time domain, i.e. directly on the speech waveform (e.g. zero crossing analysis), and those which work in the frequency domain, i.e. on the spectrum of the signal (e.g. inverse filtering). Neither of these techniques is entirely unproblematic since there are many complicating factors. It will suffice to give two examples. Time domain techniques for instance operate on the assumption that the excitation source is regular. Lieberman (1961) provides convincing evidence that in normal phonation, speech is not entirely regular. In the frequency domain, it is often the case that F0 coincides with the first formant of the signal, esp. in female voices, with the result that automatic pitch determination is complicated.

4. The Psycho-acoustic Nature of Speech Melody.

From a perceptual point of view, the correlate of speech melody is pitch, which can be conveniently defined as the perceptual response to the frequency pattern of auditory signals. In many simple signals, pitch varies monotonically with frequency. Nevertheless, it should not be regarded as a synonym of fundamental frequency. As a first approximation, it can be said that pitch and F0 maintain a very close relationship.

There are different scales available to relate F0 to pitch. One scale derives from the musical scale and it treats all logarithmically equal frequency differences as being equal pitch differences. Experimental research has shown that this equal-ratio division does not express the relation between pitch and frequency accurately. Stevens (1937) reports an attempt to establish this relationship by means of formal listening experiments, in which subjects were asked to control F0 dials until a pitch sounded, for instance, half as high as a particular standard. This has resulted in the mel-scale (Stevens & Volkman, 1940), which deviates from the musical scale, esp. in the lower and higher frequency areas (figure 1.8):

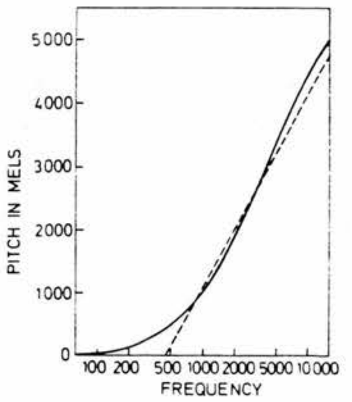


Figure 1.8: The relation of pitch to frequency as determined by judgement of tonal distances. The dashed line indicates equal numbers of mels per octave. In the frequency range most frequently used for speech melody the mel curve departs most from a straight line. (Adapted from Stevens et al., 1937).

4.1. Pitch and Sinusoidal Signals.

Pitch scales such as those described above indicate that changes in the F0 of a tone are paralleled by changes along the perceptual dimension of pitch, i.e. different frequencies lead to different pitches. A fundamental question in this respect concerns the smallest frequency differences which are still perceivable as differences in pitch. This question is addressed in Shower & Biddulph (1931). They used frequency-modulated tones to establish the just noticeable difference (JND). In the experiment, subjects were presented with pure tones which stayed at a single frequency for a time, after which they moved to another frequency. Each time the frequency movement became progressively smaller. This technique was applied to investigate the perception of frequency change in various frequency areas and at different levels of intensity. The results of this experiment are given in figure 1.9:

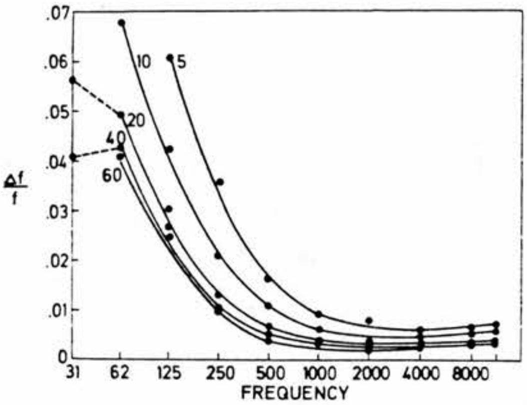


Figure 1.9: The relative change in frequency (Δ f/f) necessary to hear a change in pitch. This is the amount of change necessary when the tone changes slowly and continuously. Numbers at the top left of each curve denote the sensation level of the tone (From Shower & Biddulph, 1931).

It can be seen quite clearly that the relative change in frequency necessary to hear a change in pitch does not remain constant throughout the whole frequency range: the JND gets progressively smaller in the lower frequency area and levels off at about 1000 Hz. Moreover, it can be observed that generally a larger change is needed at lower intensity levels.

More recent attempts to establish the JND of frequency change are reported in Wier, Yestead & Green (1977). Their experimental design was different from the above, in that their stimuli consisted of pulsed sinusoidal signals. These tones were presented sequentially to subjects in a pairwise fashion, and on each trial, subjects were required to indicate which tone was the higher of the two. The experiment started with a stimulus difference which was well above the threshold. During the session, two consecutive correct judgements led to a decrease of the stimulus difference, whereas an incorrect response led to an increase. The collected data are summarized in figure 1.10:

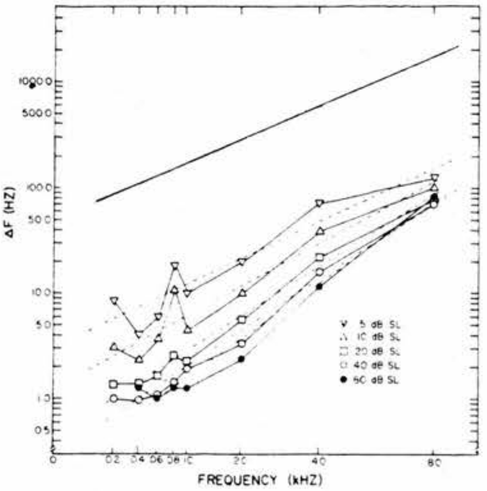


Figure 1.10: Mean JND for different intensity levels averaged across subjects (Adapted from Wier et al., 1977).

As in Shower and Biddulph's experiment, a significant relationship was found between frequency discrimination and the frequency range, as well as with the sensation level of the presented tones. When these results and those of various other experiments are compared to those of Shower and Biddulph (figure 1.11), it can be seen that their estimates of the JND are considerably larger for low frequencies, about equal at 2000 Hz and smaller at frequencies above 2000 Hz.

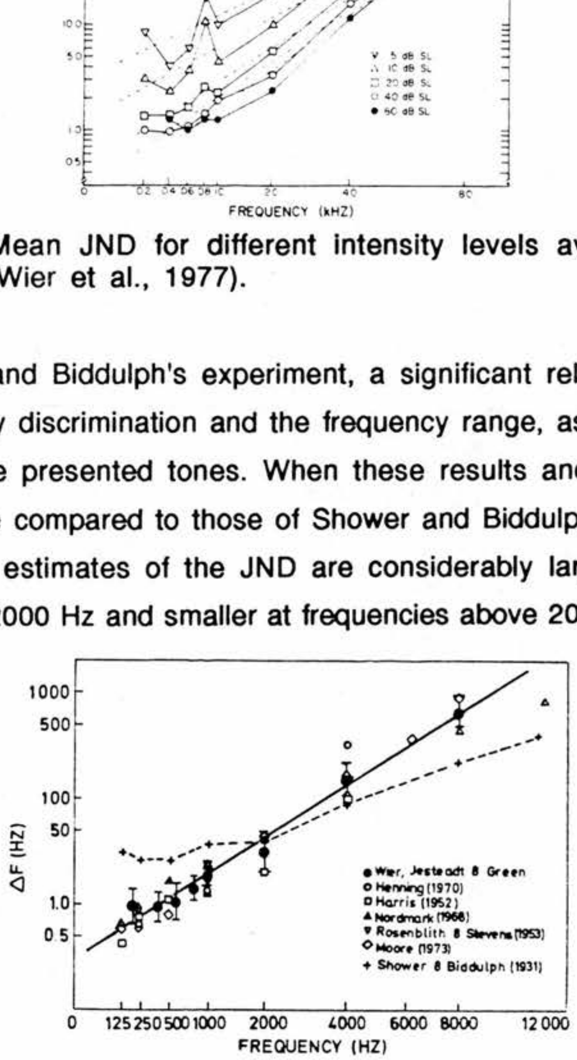


Figure 1.11: Data and function for the stimuli of 40 dB. The data of the other authors are for stimulus levels between 30 and 50 dB. All the data are for pulsed sinusoids except Shower and Biddulph (Adapted from Wier et al., 1977).

These differences may just be an artifact of the experimental method used for establishing the JND and the different kinds of stimuli involved (modulated vs. pulsed sinusoids). The interest in these data is however not the precise value of the JND, but rather the fact that frequency discrimination for pure tones is extremely good: in the 200-400 Hz range, subjects reliably discriminate frequency differences of as small as 1 Hz at higher intensity levels.

Another matter of importance in frequency discrimination is the relationship between the duration of a tone and the perception of pitch. All the experiments discussed so far have used tones of fairly long duration. Moore (1973) reports results for tones of much shorter duration down to as short as 6.25 msec. Even for these very short tones, good subjects can hear frequency variations of less than 1% for frequencies of 1000-4000 Hz and less than 2 % for 500 Hz. These results suggest that shorter and shorter tone bursts continue to elicit an identifiable pitch response, the variability getting progressively larger as the duration decreases.

4.2. Pitch and Complex Signals.

By far the majority of signals people are confronted with in everyday life are by nature not sinusoidal but complex. This is most certainly true for speech signals which are never sinusoidal. In this section, the relationship between F0 in such complex signals and pitch perception is briefly examined in fairly general terms. Subsequently, a detailed discussion of frequency discrimination and various aspects of pitch perception in speech-like signals is presented.

In a most elementary perspective, pitch is perceived in complex signals on the condition that the higher frequencies are harmonically related to the fundamental frequency, i.e. when they are whole-number multiples of the fundamental. The physical presence of the fundamental itself is no essential requirement for pitch to be perceivable. This can easily be illustrated by band-limited telephone speech, in which the frequency range of the speech signal is limited to between 300 and 3500 Hz. In the case of a male native speaker with a low fundamental frequency of around 100 Hz, the fundamental component is absent from the telephone signal. Nevertheless, speech melody variations are clearly perceivable, which suggests that the presence of the fundamental is not required for pitch to be perceivable.

The reason for this is that the value of the fundamental can be derived from the relationship between the harmonics in the speech signal. If a sound for instance has harmonics at 300, 400, 500 and 600 Hz, the fundamental must be 100 Hz. Even though the fundamental is not physically present, the waveform is periodic and has a frequency of 100 Hz, as indicated in figure 1.12:

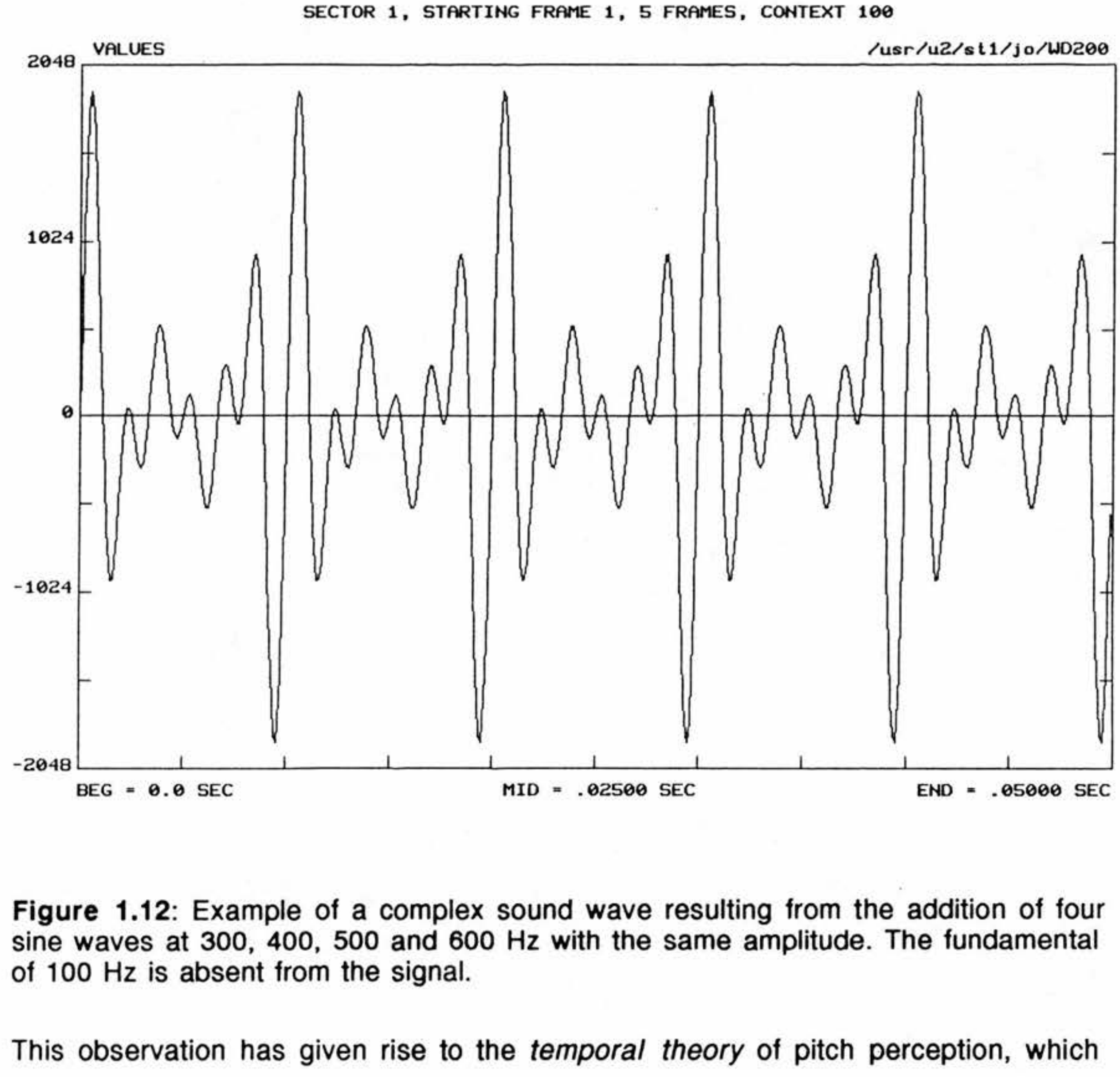


Figure 1.12: Example of a complex sound wave resulting from the addition of four sine waves at 300, 400, 500 and 600 Hz with the same amplitude. The fundamental of 100 Hz is absent from the signal.

This observation has given rise to the *temporal theory* of pitch perception, which assumes that the human ear determines the pitch of a complex signal by measuring the periodicity with which the cycles in a wave repeat themselves in time. This theory of pitch perception gives accurate pitch estimates in most instances, but there are several problems. It has, for instance, problems in accounting for ambiguities in the perception of pitch. In a complex tone with harmonics at 1910, 2110 and 2310 Hz, the perceived pitch is ambiguous as either 192 Hz or 211 Hz, since within the resolution of the human ear, the harmonics can be interpreted as whole-number multiples of either of these frequencies.

Observations of this kind have suggested a theory of pitch perception in which the basilar membrane of the ear carries out a frequency analysis of the sound. Only the frequencies of those components in the sound which are strong enough to be detected are registered. A pattern recognition procedure consequently establishes the fundamental as the best fit to this range of frequencies. This theory of pitch perception is known as the *place theory*. Generally, it can be said that the lower harmonics contribute to a larger extent to the perception of pitch than the higher harmonics, which can often not be distinguished from each other because their representation on the basilar membrane is insufficient. As such, they cannot contribute to the perception of pitch. The most important contribution is made by the frequency area in which the harmonics are clearly separable, i.e. the spectral area between 300 and 2000 Hz.

4.3. Frequency Discrimination.

The discrimination of steady-state frequency levels in speech-like signals was investigated by Flanagan and Saslow (1958) and Klatt (1973). Flanagan and Saslow (1958) synthesized four vowels [i, u, ɑ, ɛ] by means of a cascade formant synthesizer, each vowel having a 500 msec duration. They were provided with a standard F0 level of either 80 or 120 Hz at intensities of 60, 70 or 80 dB. From these standards, the stimuli were derived by raising or lowering the frequency by 0.0, 0.3, 0.6, 0.9, 1.2 and 1.5 Hz. Standards and stimuli were combined into AX discrimination items and presented to informants who were asked to judge whether the second vowel was higher or lower in pitch than the first.

The JND, which was defined as the frequency change which elicits correct responses 75 % of the time, was established at 0.30 Hz. Flanagan and Saslow (1958) furthermore conclude that:

> The discrimination data for the several vowel conditions exhibit no particularly impressive variations with vowel, sound pressure level and fundamental frequency. (p. 440).

The only significant trend in the data is that the JND is smaller at lower intensity levels, which is contrary to the observations made for sinusoidal signals. This is accounted for by the fact that the higher frequency components in the stimuli are masked by the lower ones.

Klatt (1973) also carried out a number of experiments to investigate the JND of steady-state frequency levels in vowel-like stimuli. He synthesized /ɛ/ and /ya/ stimuli in which the F0 was manipulated as an experimental variable. The stimuli were presented in AX format and subjects were required to judge whether stimulus X was higher or lower in pitch than stimulus A. Frequency discrimination was found to be extremely accurate, in that a JND of 0.3 Hz was found for /ɛ/ stimuli and 0.5 HZ for /ya/ stimuli. These results agree well with those in Flanagan and Saslow. They furthermore suggest that the presence of vowel transitions does not have a significant effect on frequency discrimination.

Fundamental frequency in actual speech is rarely -if ever- stationary. It is therefore important to get an insight into the perception of frequency transitions. This point is also addressed in Klatt (1973) in a variety of experimental conditions using the same vowel-like stimuli described above. In one condition, the stimuli were assigned frequency falls and presented for discrimination in AX format. An important characteristic of these stimuli was that the average F0 in two stimuli of a discrimination item was different. The results of this experiment suggest that discrimination of these falling F0 transitions is less finely tuned: the JND was established at 2 Hz and 2.5 Hz for /ɛ/ and /ya/ stimuli respectively. Consequently, it can be argued that the type of F0 transition determines the accuracy of discrimination to a appreciable degree, in that real F0 transitions are not discriminated as well as steady-state F0 levels.

Klatt (1973) made a similar observation in an experiment in which the slope of F0 transitions was manipulated while keeping their average F0 constant. Two types of slopes were used: one in which the transition was fairly level (12 Hz/sec), while in the other the transition was rather steep (32 Hz/sec). These stimuli were presented for discrimination in an ABX format, in which subjects had to judge whether X was the same as A or as B. The JND in the level condition was found to be 1.5 Hz, whereas it was 4 Hz in the steep condition. From this, Klatt concludes that discrimination becomes worse as the slope of an F0 transition becomes steeper.

The latter results are compatible with those reported in Nabelek and Hirsch (1969), who investigated discrimination of F0 transitions in sinusoidal signals. One of the experimental conditions focused on the frequency area of around 250 Hz, in which transitions were used which either started or ended at 250 Hz, with frequency drops or rises of 28, 83 or 125 Hz. The slopes of these transitions were manipulated by varying their duration. Consequently, reference stimuli with transition durations of 10, 30, 100 and 300 msec were obtained. These stimuli were paired with test stimuli having the same or a different slope and were presented for discrimination in an AX-format. The results of this experiment indicate clearly that the JND is largest for the reference stimuli with the steepest slopes.

In conclusion, it can be said that frequency discrimination in complex, speech-like signals compares quite well with that of sinusoidal signals, especially concerning the discrimination of F0 levels. The discrimination of F0 transitions is clearly dependent on the slope of the transition. This seems to indicate that discrimination is in some way related to time: good discrimination is only possible when the frequency remains fairly stable over longer periods of time. This observation agrees well with Moore (1973), who showed that F0 discrimination is less consistent in shorter signal durations.

4.4. Pitch of F0 Transitions.

Regarding the relationship between F0 transitions and the perception of pitch it is necessary to know whether F0 transitions are perceived as unambiguous pitch levels or in terms of their start or endpoints. A very informative study in this respect is Nabelek et al. (1970), which reports a series of experiments in which informants were asked to adjust the F0 level of a steady-state frequency tone in such a way that its pitch matched that of a series of test tones containing a linear frequency change. The transitions were either rising or falling and the frequency characteristics of the test tones were manipulated as the independent variable.

In one set of stimuli, the frequency change took place over the total duration of the sound. The results suggest that informants perceived the change as an unambiguous pitch, provided that the change in frequency is fairly small (less than 50 Hz) and the duration of the frequency transition is not too large. In this condition, informants consistently adjusted the frequency of the steady-state tone to a frequency near the middle of the transition in the stimuli. This tendency was slightly more outspoken for rises than for falls. For larger excursions over longer durations, there was less consistency in the pitch matches, indicating that informants perceived the transitions as glides from one pitch level to another one.

In the second set of stimuli, the frequency transition lasted only 25 % of the total stimulus duration. In addition to this, the stimuli were constructed in such a way that this transition was located at different points in the sound, so that it was followed and/or preceded by a steady-state frequency. The structure of these stimuli is given in figure 1.13:

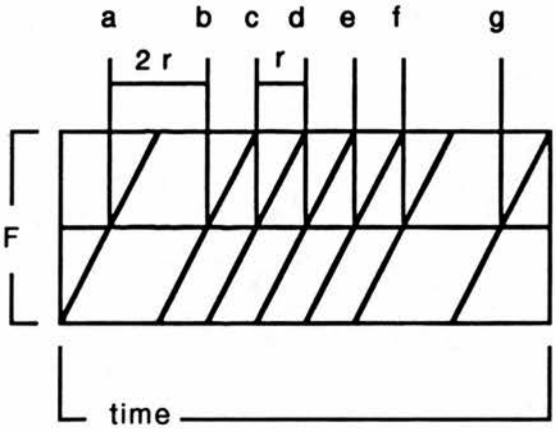


Figure 1.13: Temporal positions of the rising frequency transitions for the tone with linear frequency change during 25% of the total duration of the tone. a, b, c, d, e, f and g are the positions of the transition. r is a unit of delay of the transition used in the experiments (Adapted from Nabelek et al., 1970).

With stimuli of this kind, three basic response patterns are observed, which are the same for both stimuli with rising and falling movements. In the stimuli with small frequency excursions (50 Hz), informants heard an unambiguous pitch, the height of which correlated strongly with the location of the transition in the stimulus. This pitch level generally corresponded to the average of the frequencies in the stimuli. In the larger frequency transitions (200 Hz), the subjects were indecisive about the corresponding pitch level, in the sense that pitch matches were dispersed over the whole range of frequencies in the stimuli. In large F0 changes (580 Hz), the adjustments clustered round the initial or final frequency, irrespective of the location of the transition. This suggests that informants heard two pitches.

It should be borne in mind that the stimuli in above-mentioned experiments are in most cases not representative of F0 variation in speech: the frequency range under investigation was situated between 400 and 1500 Hz. Moreover, the excursion sizes that were used (50-580 Hz) were larger than those typical in speech. Finally, the stimuli were sinusoidal. Nevertheless, the results are in close agreement with those reported in Rossi (1971, 1978), who used more speech-like stimuli and F0 variations.

In Rossi (1971, 1978), a number of experiments are reported in which a natural vowel of [a] quality was used as the basis for the stimuli. Factors such as vowel duration (100-300 msec), frequency range (male vs. female voice) and phonetic context (isolated vs. embedded) were carefully controlled. The transition took place over the whole duration of the vowel. Rossi's primary aim was to establish the discrimination threshold for rising pitch movements, i.e. he aimed to establish the minimal pitch change necessary for informants to notice the movement as a change in pitch (cf. section 4.6. of this chapter). In one subexperiment however, subjects were presented with pairs of vowels, the first of which contained an F0 transition, whereas in the second F0 was steady. The subjects were asked to judge whether the steady-state vowel was higher, equal or lower in pitch than the vowel with the transition. It was found that subjects matched the transition with a pitch level, the frequency of which corresponded to that of the middle third of the transition. No significant difference was found between rising and falling movements.

In summary, the results of above-mentioned experiments seem to suggest that F0 transitions are perceived as unambiguous pitch levels corresponding to a pitch somewhere near the middle of the transition, provided that the transitions are small and relatively short. Furthermore, there is no indication that subjects respond differentially to rises and falls. In pitch transitions which last only a fraction of the duration of the stimulus, perceived pitch depends on the location of the frequency change, that is: on the proportion between high and low steady state pitch preceding and following the transition.

It should be emphasized that the findings reported in these studies may not reflect what happens in reality. Given that informants were only provided with a set of steady state reference stimuli, they could only make choices among these. This, however, does not necessarily mean that they actually do perceive the transition as an unambiguous pitch. This is indicated by the fact that these results seem to be incompatible with those in Klatt (1973). It is indeed a fundamental characteristic of some of Klatt's stimuli that their overall F0 levels are identical. This means that the average F0 level in the middle third portion of the stimuli is roughly equal. If transitions are indeed entirely perceived in terms of average pitch levels, as suggested above, informants in Klatt's experiment should not have been able to report any perceptual differences. They nevertheless did hear differences, which suggests that the findings in both sets of experiments are difficult to reconcile.

4.5. Excursion Size.

The question of the sensitivity of informants to pitch excursions size was addressed in 't Hart (1981). In his experiment, subjects were presented with pairs of stimuli, each containing a pitch movement in the same direction of variable size. It was the task of the subjects to indicate the stimulus with the larger movement.

The design of the experiment and stimuli was rather complex. The pitch range covered by the excursions in the stimuli varied from 1 to 6 semi-tones and the stimulus pairs contained no fixed reference: the initial and/or final frequency was different in both stimuli of a discrimination item. In order to avoid extra cues of excursion slope and duration, the rises started in the prevocalic voiceless consonant of the experimental syllable and the falls ended in the voiceless postvocalic consonant. In addition, it should be mentioned that the overall pitch contours of the stimuli were situated in three registers, depending on their initial frequency (115 Hz, 135 Hz and 160 Hz).

The results obtained in this experiment suggest that excursions of falls were more difficult to judge than those of rises. In the fall condition, 61 % of the participating subjects were found to be non-discriminators, i.e. they were unable to discriminate differences of less than 4 semi-tones. The number of non-discriminators in the rise condition was only 16 %. The discriminators needed 1.8 semi-tones to judge rises as having a different excursion, whereas the threshold for a fall was slightly larger (2.4 semi-tones). 't Hart suggests that this differential sensitivity can be explained by the fact that native speakers of Dutch are more frequently exposed to rises, since the Dutch intonational system prefers rising movements for realizing pitch accents.

't Hart's experiment is essentially non-linguistic in that the subjects had to make pure psycho-acoustic judgements. The results of more recent linguistically oriented work suggest that much smaller differences in excursion size correlate significantly with prominence judgements. In Rietveld & Gussenhoven (1985), subjects were presented with Dutch sentences containing two accents, which were realized with a rise, immediately followed by a fall. The peak height of one of the accents was manipulated as a variable, and subjects had to judge the prominence relationship between the accents. Results indicate that differences of 1.5 semi-tones correlate reliably with differences in prominence judgements. A prominence rating experiment in Verhoeven (1986) indicates that differences of as small as 1 semi-tone are correlated with different prominence ratings. Although the results of all these experiments are not necessarily comparable, they suggest that the thresholds in 't

Hart may be on the large side. They are furthermore indicative of the essential difficulty in mapping psycho-acoustic variables onto linguistic ones, a point which will be discussed further in the next chapter.

4.6. The JND of F0 Change.

For work in this area, we have to come back to Rossi (1971, 1978). In Rossi (1971), the perception of rises was investigated in a variety of experimental conditions on both natural and synthetic vowels. Informants were presented with a series of vowels containing rises of various sizes. The stimuli were presented one at the time and it was the task of the subjects to indicate whether they heard a rising tone or not.

Rossi draws two basic conclusions from his results. On the one hand, it can be said that the threshold for a movement to be perceived as a transition depends on its duration. That is to say that the threshold is largest in transitions of short duration and smallest in durations up to 200 msec. This is summarized in table 1.2:

Duration	Threshold
50	4.93
100	2.86
200	2.27
300	-

Table 1.1: Summary of the different thresholds as a function of stimulus duration. The thresholds are in semi-tones.

For durations longer than 200 msec, the threshold seems to stabilize. On the other hand, the threshold seems to be related to the frequency of the transition onset or the pitch register under investigation:

> Le seuil est relativement plus élevé pour une voix de femme que pour une voix d'homme (Rossi, 1971: 9).

This is indeed true if the threshold is defined on the Hz scale: in those terms the low register has a threshold of 19 Hz, whereas that of the high register amounts to 35 Hz. On a logarithmic scale however much of the difference is eliminated since the threshold is 2.27 semi-tones in the low register and 2.79 semi-tones in the high register.

5. Summary.

In this chapter, a working definition of intonation has been proposed in which this linguistic concept is related to speech melody. Subsequently, the physiological, acoustic and psycho-acoustic nature of speech melody has been investigated.

In a physiological perspective, speech melody is correlated with rate of vocal fold vibration, which is accounted for by the aerodynamic-myoelastic theory. In this theory, phonation is related to the aerodynamic properties of the airstream and the myoelastic characteristics of the laryngeal system. The rate of vocal fold vibration in phonation is shown to be determined by the primary factors of vocal fold length (tension), which is finely regulated by activity of the instrinsic laryngeal musculature, and subglottal pressure. Furthermore, a number of secondary factors are described such as coupling effects of the larynx with the sub- and supraglottal cavities. As far as the control mechanisms of speech melody are concerned, it is observed that different aspects have to be attributed to different physiological mechanisms. The overall decline of F0 in intonation contours is primarily related to subglottal pressure, whereas the localized deviations from the declination line are found to be associated with laryngeal muscle activity, esp. that of the cricothyroid.

In an acoustic perspective, speech melody is said to be manifested in the speech signal as fundamental frequency variation.

In a psycho-acoustic perspective, speech melody is related to pitch variation. First, the relationship between frequency and pitch in sinusoidal signals has been discussed. From this, it emerges that listeners are highly sensitive to very small variations in frequency. This sensitivity is determined by the frequency range (the JND is smallest in low frequency areas), the intensity of the signal (the JND is smallest at high intensity levels) and the duration of the signal (the JND is smallest in signals of longer durations).

As far as complex, more speech-like signals are concerned, it has been shown that frequency discrimination is related to the type of frequency transition involved. In the discrimination of steady-state frequencies, a high sensitivity is observed, similar to that in sinusoidal signals. This sensitivity clearly decreases with a factor of a magnitude in the discrimination of sloping frequency transitions, which are most typical in speech signals.

As to the perception of pitch of these frequency transitions, the evidence is conflicting. Some experiments indicate that informants perceive frequency transitions in terms of steady-state pitch levels, the value of which corresponds to the average frequency in the middle third of the stimuli. Other experiments do not corroborate this claim. Concerning the discrimination of pitch excursion size differences, it emerges from 't Hart (1981) that subjects are more sensitive to rising excursions (1.8 semi-tones) than to falling excursions (2.4 semi-tones). Finally, the JND of frequency change has been reported to be dependent on the duration of the transition. In transitions of short durations, the JND is rather large, i.e. 5 semi-tones. In longer durations, the JND is considerably smaller at ca. 2 semi-tones.

CHAPTER 2

Analytic Models of Dutch Intonation.

0. Introduction.

In this chapter, several models of Dutch intonation are discussed. They all have the common characteristic that they are essentially what we would like to call form-driven studies in that they have at their centre the belief that the concrete acoustic parameters in the speech signal are realizations of more abstract tonal categories: it is their aim to arrive at a formal representation of the underlying tonal organization. Functional considerations are in all instances excluded outright. The models that are discussed, however, differ in the abstractness of their formal representation of intonational structure and in their consideration of the relationship between intonation and prominence.

The first approach is associated with the researchers of the Institute of Perception Research (IPO) in Eindhoven, The Netherlands. Their model, which will be referred to as the Grammar of Dutch Intonation (henceforth GDI), reflects a major concern with speech synthesis. It is fairly concrete in the sense that it is firmly founded in phonetics. In this domain however, it is undoubtedly as abstract as it can be.

The other approaches are rather more abstract in the sense that their formal representations are phonological in nature: their aim is to provide an insight into the linguistic organization underlying phonetic observations. Thus, they try to arrive at generalizations about abstract intonational structure. For this purpose, Ladd (1983a, 1983b) proposes a peak-feature model, in which he analyzes intonation in terms of features, in very much the same way as the phonological analysis of segments. Gussenhoven (1988a, 1988b) proposes a treatment of Dutch intonation in terms of autosegmental phonology. His approach describes intonation in terms of a small number of underlying tonal morphemes from which intonation contours are derived by means of modification processes and phonological rules.

1. The Grammar of Dutch Intonation.

The model that is described in this section was developed by researchers at the Institute of Perception Research in Eindhoven during the 1970's and 1980's. Research at this institute has not only resulted in a very comprehensive phonetic model of Dutch intonation but more recently, the IPO analysis techniques have been applied successfully to English (Willems, 1982, 1983, Willems et al., 1988, De Pijper, 1984 and Collier 1989), German (Adriaens, 1984) and Russian (Odé, 1989). It has nevertheless received little widespread recognition. This is unfortunate, since the techniques for intonation analysis are innovative and original, yielding results which are experimentally verifiable. The model will therefore be discussed here in a fair amount of detail.

1.1. Basic Assumptions.

The most fundamental assumption of the IPO approach to the analysis of intonation is formulated in Cohen & 't Hart (1967) and 't Hart (1984):

> The F0 curve is a superposition of pitch movements relevant to the perception of speech melody, and of other variations, that are merely ascribable to the irregularities in the oscillation behaviour of the vocal cords ('t Hart, 1984: 195).

Furthermore, at the level of speech production it is argued that:

> (...) it is unthinkable that all the variations observed are the results [sic] of an active neuro-muscular control. Consequently, it would be useless for a speaker to have any conscious intentions to produce them ('t Hart, 1984: 194).

In the production stage, a distinction is thus assumed between *voluntary* and *involuntary* changes in the periodicity of vocal fold vibration. Voluntary changes result from variations in the states of laryngeal musculature, insofar as they are actively controlled by a speaker. These changes give rise to gross F0-variations in utterances (macro-intonation). Involuntary changes on the other hand fall beyond the speaker's control, since they are determined by other factors. They are primarily:

> (...) dependent on the air pressure drop across the glottis, and on the accidental height of the larynx, on the position of the tongue and of many other

> articulators. Therefore, they are dependent on segmental, phoneme-sized aspects ('t Hart, 1979a: 368).

All these segmental effects are responsible for local perturbations in the F0-curve (micro-intonation).

Given these assumptions, an analysis of intonation can attempt to separate any F0 curve into F0 events which are produced voluntarily and those that are not produced voluntarily, an aim which can be achieved by extracting the discrete commands to the laryngeal musculature in the production of speech melody.

In a perceptual dimension, it is furthermore assumed that only voluntarily produced F0 variations are perceivable:

> (...) it is counter-intuitive that any speaker intends to produce sounds and sound variations that might not be perceptible ('t Hart, 1984: 195).

This assumption makes it possible to analyze F0 curves in terms of F0 variations that are perceivable and those that are not. Only the F0 variations that are produced voluntarily can be expected to be perceivable and relevant to intonation. It is this type of analysis that was conducted in the IPO approach and it was achieved by means of the stylization method.

1.2. The Stylization Method.

In order to investigate the F0-variations relevant to the perception of intonation, the *stylization method* was developed. Basically, it involves the construction of perceptual analogs to F0 curves: i.e. the actually observed F0-curves in utterances are replaced by artificial contours which are perceptually indistinguishable from the original ones. This was originally achieved by means of the intonator, an instrument which consists of a bank of band-pass filters measuring the spectral composition of the input speech signal. These measurements were used to resynthesize the utterance, F0 being externally controlled by the experimenter.

In order to relate F0-changes in utterances to perception, two comparative procedures were used, both of which attempt to obtain perceptual equivalence between the original contour and the artificial stylization. These methods are *close-copy stylization* and *analytic listening*.

A close-copy stylization is defined by 't Hart (1986) as:

> A piecewise linear approximation of an F0 curve with the smallest number of straight line segments that meets the requirement that the synthesized version with this contour is audibly indistinguishable from a resynthesis of the original (p. 1838).

A close-copy stylization consequently has to meet two criteria: it should contain the smallest number of straight lines and at the same time be perceptually identical to the original contour.

In practice, a close copy stylization is obtained like this: the sentence under analysis is recorded on the loop of a tape. F0 is measured and displayed on a screen. The experimenter now attempts to make the best possible copy of the original contour. Once this has been achieved, he tries to leave out as much detail as possible, continuously comparing auditorily how the contour sounded before and after removal of the detail. In experimental practice, this involves smoothing out F0 changes, i.e. the F0-curve is replaced by straight lines. By doing this, an attempt is made to find the minimal specification for the artificial contour, which still yields the same perceptual impression as the full specification.

In analytic listening, the F0 measurement stage is omitted. The experimenter concentrates on the melodic aspect of the utterance and tries to abstract away from the verbal information. He attempts to convert the results of his analytic listening directly into F0 commands to the intonator.

The result of the stylization method is that the complex F0-curves of utterances present themselves to the experimenter as a sequence of discrete F0 events, the combination of which gives the same melodic impression as the original contours. These events are called 'perceptually relevant pitch movements' and are represented in the stylization as straight line segments (figure 2.1):

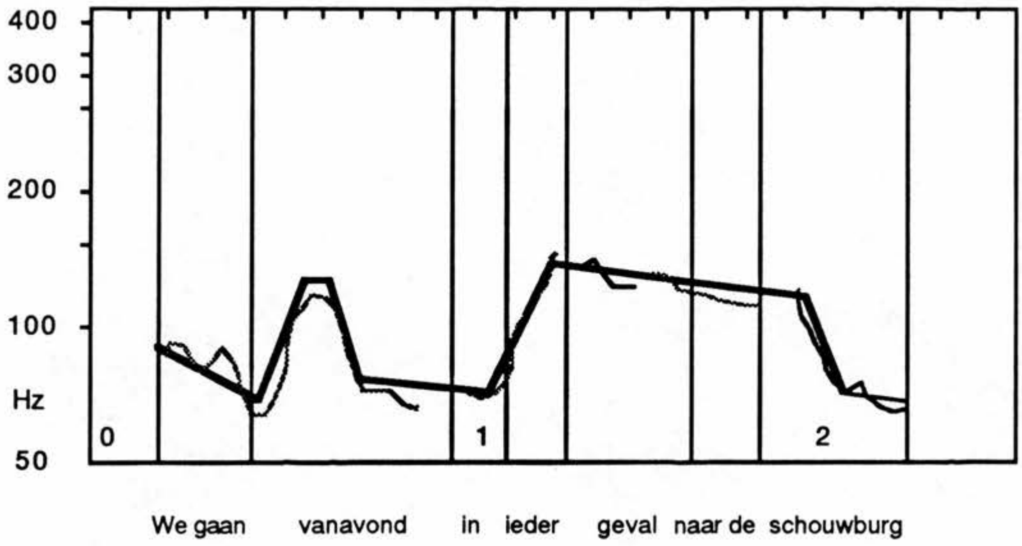


Figure 2.1: Example of a fundamental frequency curve (dotted line) with stylized pitch contour that is indistinguishable from the original (Adapted from Willems, 1983: 40).

As the first stage in the analysis of Dutch intonation, the stylization method was applied to a large corpus of various speech materials, such as spoken news bulletins, radio commentaries, spontaneous conversations etc.

1.3. The Standardization of Pitch Movements.

The next step in the analysis consists of *standardizing* the pitch movements with respect to their acoustic characteristics. This was achieved by a series of experiments into the tolerance levels in the perception of these acoustic variables by native speakers of Dutch. Some of these experiments have already been discussed in the previous chapter, but for convenience sake, the main results are briefly summarized again.

As far as the *speed of change* of pitch movements is concerned, the GDI distinguishes between abrupt and gradual pitch changes. The results of psycho-acoustic experimentation summarized in 't Hart (1976) indicates that different slopes are likely to be distinguished only if the duration of the movement is longer than 250 msec. Since the duration of an average-sized syllable is usually shorter than 250 msec, it is concluded that differences in slope of pitch movements are not perceived in the syllable domain. It is only when the pitch changes gradually over several syllables that differences in slope may become noticeable. This has resulted in the view that

only declination and inclination are potential candidates for exploiting differences of slope perceptually, since both extend over several syllables (Collier, 1983).
The second variable in the standardization of pitch movements concerns the parameter of *excursion size*, which is highly variable in actual speech. An attempt was therefore made to determine the sensitivity of the human ear to pitch distance differences in 't Hart (1981). It was found that listeners need 3 to 4 semi-tones to distinguish differences in excursion size with a reasonable degree of certainty. Van Katwijk (1974) and more recently Gussenhoven & Blom (1978) provide evidence that excursion size correlates significantly with the perception of syllable prominence.
Finally, there is the parameter of *timing* of pitch movements. This concerns the location of a movement with respect to the vowel onset of the syllable. In the GDI, a three-way distinction is made between early, late and very late pitch movements. Support for this distinction is found in an experiment by Collier (1975a), the results of which indicate that informants classified a pitch contour into three categories depending on the location of the pitch movement in the experimental syllable. Like pitch excursion size, the acoustic dimension of alignment has been shown to play a role in the perception of prominence. This is suggested by research reported in Govaert & Van Katwijk (1968), who carried out an experiment, in which the location of pitch movements was shifted through a nonsense utterance. These stimuli were presented to Dutch informants for prominence judgement. The results show that the alignment of pitch movements is a strong cue to the perception of prominence, if a rise is situated early in a syllable. A fall on the other hand has to be located typically late in the syllable. The results of Van Katwijk & Govaert are corroborated by one of Collier's experiments (Collier, 1970), in which informants were asked to adjust the location of a pitch rise in a syllable in such a way that the syllable was perceived as most prominent. In all instances, the rises were positioned early.
By means of experimentation of this kind, the great variability in the pitch movements observed in the stylizations was standardized with respect to a small number of perceptual dimensions: *direction* (rise vs. fall), *speed* (abrupt vs. gradual), *timing* (early, late or very late in the syllable) and *size* (small, medium or large). Furthermore, pitch movements also correlate with a linguistic dimension, i.e. that of *prominence*: some pitch movements are prominence-lending, whereas others are not. This aspect refers to the relationship between intonation

and prominence, which is discussed in detail in section 1.8 of this chapter. The standard values of the pitch movements in Dutch are summarized in table 2.1:

Type	Prominence lending	Place in Syllable	Duration (msec)	Excursion (semi-tones)
RISE				
1	+	Early	100	+4
2	-	Very late	100	+4
3	+	Late	150	+6
4	+ or -	Duration depends on accent interval		
5	+	Before falls A or D	50	+2
FALL				
A	+	Late	75	-5
B	-	Early	75	-5
C	-	Very late	20-50	Undefined
D	+ or -	Duration depends on accent interval		
E	+	Early	38	-2.5
F	+	Succession of E falls		

Table 2.1: Standard values for perceptually relevant pitch movements in Dutch (From Van Geel, 1983: 8).

1.4. The Reality of Pitch Movements.

Evidence for the reality of pitch movements as discrete units in intonation has been provided from the physiology of speech production. As pointed out in chapter 1, Collier (1975b) found that rises and falls correlate with discrete commands to the laryngeal musculature, esp. the cricothyroid muscle. EMG measurements reveal a straightforward relationship between activity of this muscle and the size of the pitch excursion (large vs. small), direction of the pitch movement (rise vs. fall) as well as to its slope (steep vs. gradual). Alignment differences were however not investigated systematically. These findings confirm the basic assumption that perceptually relevant pitch movements are under voluntary muscle control of the speaker, and hence have a reality as units in intonation.

In passing, the reader's attention is drawn to the fact that in this respect there is a good analogy between how pitch movements in speech are produced physiologically and the instrumental method in the IPO approach for discovering them.

Physiologically speaking, F0 is controlled by neural commands to the laryngeal musculature. In the stylization process, perceived pitch (in analytic listening) controls the commands which are given to the intonator by the experimenter, in order to build a perceptually equivalent copy of the pitch pattern.

1.5. Declination.

The phenomenon of declination deserves separate attention, since it is clearly regarded as a perceptually relevant dimension of Dutch intonation. It is clear that stylization of an utterance without declination has a perceivable effect (Collier, 1983), but in the early stages in the development of the GDI, the phenomenon only received marginal attention.

Declination is defined by Pierrehumbert (1979) as 'the tendency of pitch to drift downwards over the course of an intonation group' (363). The result of this is that declarative utterances have a lower overall pitch at the end than at the beginning, as shown in figure 2.2:

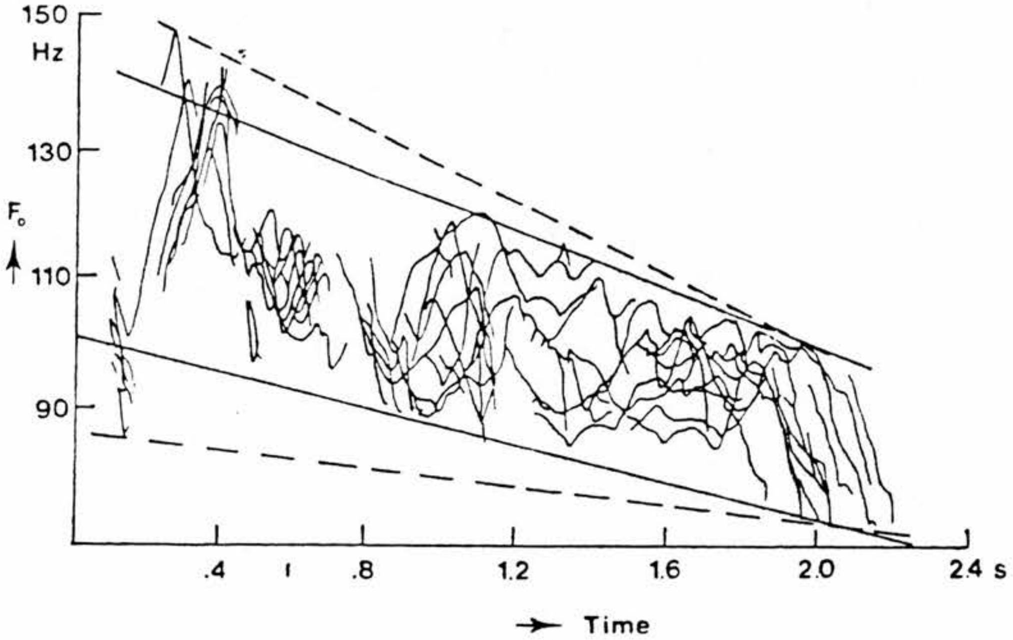


Figure 2.2: Superposition of 11 different utterances of equal length by one speaker. The dashed lines represent best fits through the absolute F0 maxima and minima; the solid lines are the fits through the majority of maxima and minima (Adapted from Cohen et al., 1982).

In the GDI, declination was originally regarded as a purely operational concept, which enables the model to relate pitch movements to the average voice range of a speaker. In several other views, it is a rather more theoretical concept, which is assumed to correlate with utterance duration, syntactic phrasing and sentence type. The latter view is clearly expressed in Thorsen (1988), who argues that declination is dependent on the sentence category. In a statement, the slope of the declination line is rather steep. In a continuation and a syntactically marked question, the rate of declination is less steep. There is no declination in syntactically unmarked questions. For Thorsen's observations to be relevant, it must be assumed that declination can be actively controlled by the speaker.

Since a full account of declination is largely beyond the scope of the present work, the main research trends are only summarized, while the solutions proposed in the GDI are elaborated in detail.

1.5.1. Acoustic Characteristics.

In many descriptions of intonation, the declination line is used as a reference on which the main pitch movements (or targets) are located. This has led to the reference-line debate, i.e. the controversy over which reference line to take as the most basic one: the top line (upper declination line) or the bottom line (lower declination line). Several points of view are represented here. Maeda (1976) clearly favours the bottom-line approach, whereas others opt for the top-line view (Cooper & Sorensen, 1981), the main argument being that the two lines tend to converge: i.e. the top line declines at a higher rate than the bottom line. Therefore, both lines require independent specification.

The GDI takes the view that the specification of one line is sufficient, because in most cases the convergence of the two lines is negligible on a logarithmic scale: they tend to run parallel, so that the top line can be derived from the baseline. Cohen et al. (1982) present several arguments in support of this bottom-line approach. From an acoustic angle, the bottom line seems to be more invariant than the top line. Comparison of F0-curves has indicated that in declarative utterances of similar length, the bottom lines overlap substantially. The top line on the other hand seems to vary with respect to the excursion size of the rises. This leads to the rationale that the bottom line is a more stable reference line than the top line. Secondly, a practical consideration must be taken into account in arguing for the bottom line. It is impossible to draw a top line for an utterance containing only one peak. The bottom line, however, is even in those circumstances retrievable. A perceptual argument for the bottom line is that listeners seem to use it as a reference line, in order to determine whether a fall is terminal or whether it suggests continuation of the utterance. In the terminal fall, the movement falls until it reaches the bottom declination line. In continuation, the F0 reset is not complete.

A second point of controversy regarding the acoustic characteristics of declination is the slope of the bottom/top line. The GDI makes slope dependent on utterance length, as expressed in the following formula ('t Hart et al., 1981):

$$D1 = -8.5 \text{ semi-tones}/t \qquad t > 5 \text{ sec}$$

$$D2 = \frac{-11}{(t + 1.5)} \qquad t < 5 \text{ sec}$$

These measures were established by looking at a great number of F0-traces of actual speech utterances, and they are reported to work well in practice.

1.5.2. Physiology.

As to the physiological causes of declination it was already mentioned in chapter 1 that several hypotheses have been put forward. One holds that the declination effect is caused by a gradual decrease in subglottal air pressure during the production of utterances: the gradual drop of P_{sg} causes the vocal folds to vibrate at a progressively slower rate, resulting in a gradual drop of absolute pitch. It is however a matter of controversy whether this decrease in P_{sg} affects the rate of vocal fold vibration directly or indirectly. Research reported in Collier (1975b) and Gelfer et al. (1987) suggests that declination can be directly accounted for on the basis of falling subglottal pressure. Collier observes that the rate of F0-declination correlated well with the rate of P_{sg} decrease, since the P_{sg}/F0 ratio corresponds with the generally accepted value of 1/5.

Maeda (1976) on the other hand finds that declination is larger than can be accounted for by the estimated decrease of P_{sg}. This is assumed to result from an indirect mechanical relationship between decreasing lung volume and declination, in which tracheal pull mediates between subglottal pressure and vocal fold vibration. The collapsing lungs gradually pull down on the trachea and this tracheal pull causes a rotation of the cricoid cartilage, which results in a progressive shortening of the vocal folds. This causes a gradual relaxation of the vocal folds and consequently a lowering of absolute pitch.

Both the direct and indirect view share the belief that declination is a passive phenomenon, which is essentially a property of the respiratory system. Such a view is entirely compatible with declination as an operational concept, but not with theories which assign declination a theoretical status, since the latter assume that declination is preprogrammed and preprogramming of declination can only make sense if it can be controlled actively by the speaker.

1.5.3. Psychological Evidence.

It is also important to observe that declination has been shown to possess a certain degree of psychological reality. The results of several experiments support this, in that they indicate that listeners perceptually normalize for declination when judging the height of two consecutive pitch peaks in utterances. Pierrehumbert (1980) found that if speakers had to evaluate the height of two such peaks in the same contour, the second peak was judged to be equal to the first one when it was actually lower in terms of its absolute frequency.

A similar compensation for declination is reported by Leroy (1984). The results of her experiments suggest that listeners overestimate the height of the second peak, irrespective of whether the declination slope was standard, steeper than standard or entirely flat. She does not provide conclusive evidence as to whether listeners normalize for perceived declination or for expected (mentally computed) declination.

Secondly, Collier (1983) points out that longer contours with an objectively steady F0 are not perceived as monotonous. Instead, the utterance is perceived as having a slightly increasing pitch ('inclination'). This is confirmed in Leroy's experiment in which she used stimuli in which the peaks were located on a flat bottom-line. She found that even in these instances, subjects overestimated the second peak with respect to the first one.

1.6. Pitch Contours.

Once the discrete perceptual units of intonation are established, pitch contours can be accounted for in terms of these discrete elements. Pitch contours are defined as 'perceptual structures, composed of a string of pitch movements' (Collier, 1979: 361) or more concretely in Willems (1983: 37):

> A pitch contour consists of a limited number of stylized and standardized pitch movements which can be represented by straight lines that rise and fall between (...) parallel declination lines.

In order to describe the internal structure of pitch contours, Collier (1972) developed a grammar consisting of an explicit statement of all the permissible successions of pitch movements. This tentative grammar was subsequently checked against a corpus and it turned out that 30% of the pitch contours remained unaccounted for. Moreover, many combinations that were predicted by the grammar did not occur in the speech material. Therefore, adaptations were implemented.

Certain successions of pitch movements were found to be more coherent than others: i.e. these typical pitch clusters can be substituted as a whole by other combinations of pitch movements in the same environment. Therefore, a new descriptive unit was introduced, which occupies an intermediate level between the pitch movement and the pitch contour: the *pitch block*. The remodelled grammar then consists of a specification of the internal structure of each pitch block in terms of its pitch movements, and an explicit statement of the combinatory possibilities between pitch blocks.

The grammar distinguishes between three types of pitch blocks:

P-blocks: prefix blocks.
C-blocks: continuation blocks.
E-blocks: end blocks.

Each pitch contour then obligatorily consists of an E-block, while [P], [C] and [[P] C] are optional recursive elements which are free in number. This is summarized in the formula:

CONTOUR = [[P] C] [P] E.

The internal structure of each pitch block is schematized in fig. 2.3:

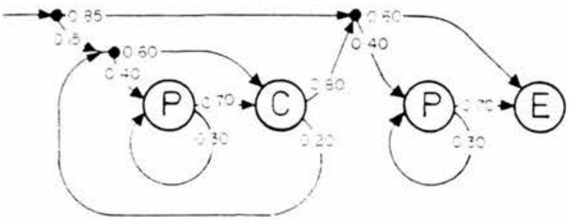


Figure 2.3: **The generation of pitch blocks. Arrows indicate the continuation possibilities after a chosen initial pitch movement. The choice among elements within square brackets is free (From 't Hart & Collier, 1975).**

In total 6 P-blocks, 8 C-blocks and the same number of E-blocks are distinguished. The internal composition of these pitch blocks is given in table 2.2:

P-blocks	**C-blocks**	**E-blocks**
P1 = 1B or 1D	C1 = 1(5)A2 or 1(5)D2	E1 = 1(5)A(2) or 1(5)D(2)
P2 = 3B or 3D	C2 = 1(5)E	E2 = 1(5)E
P3 = 4B or 4D	C3 = 1(2)	E3 = 1(2)
P4 = B or D	C4 = 2	E4 = 2
P5 = 1E	C5 = 3(2)	E5 = 3C(2) or 3D(2)
P6 = E	C6 = 4(5)A2 or 4(5)D2	E6 = 4(5)A(2) or 4(5)D(2)
	C7 = (5)A2 or (5)D2	E7 = (5)A(2) or (5)D(2)
	C8 = (5)E	E8 = (5)E

Table 2.2: Survey of the internal composition of the pitch blocks P, C and E, with a specification of their internal structure in terms of pitch movements (From 't Hart & Collier, 1975). Round brackets indicate that the pitch movement is optional. For a detailed description of the acoustic characteristics of the pitch movements, the reader is referred to table 2.1.

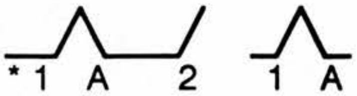
It can be noted that there is a strong resemblance between the C and E blocks: in most cases the C-blocks are identical to their E-counterpart, except for the final non-prominence lending rise which is obligatory in the former.

In a third part of the grammar, a number of ad hoc restrictions are imposed on the combinatory possibilities of the pitch blocks in general. These restrictions are:

(a) P-blocks ending in a low pitch cannot be followed by any block beginning with a high pitch. A C-block ending in high pitch cannot be followed by any block beginning with low pitch. Hence, (1) and (2) are ill-formed (1):

(1)

(2)

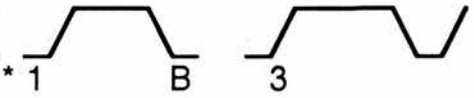


(b) Blocks C4 and E4 cannot be preceded by any P-block. Therefore, (3) is not acceptable:

(3)

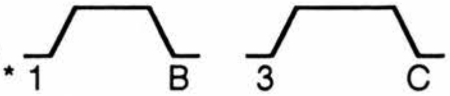
(c) The two movements in P1 and P2 cannot occur on different syllables, if they precede either C5 and C6, or E5 and E6. This excludes combinations like (4), (5), (6) and (7):

(4)



(5)

(6)



(7)

---

(1) In the examples given, the straight line segments represent the perceptually relevant pitch movements in the pitch contours. The notation underneath explicitly states the identity of the movement. In the notation, declination is not transcribed.

Verification of this version of the grammar gave correct predictions in 95% of the cases. Collier (1972) argues that the cases that remain unaccounted for are marginal and have to be considered as essentially unpredictable. It concerns instances where the speaker was interrupted before being able to produce a pitch movement or made intonation errors accompanying segmental misarticulations. Furthermore, there were a number of hybrid combinations in which the rising movements came too early or too late in the syllable as in for instance 1C and 3A.

1.7. Abstract Intonation Patterns.

In Collier (1972), intonation patterns are described as:

> (...) configurations of pitch movements that are capable of making an utterance sound intonationally complete. A pattern is a perceptual unit (Gestalt) that may extend over a variable number of syllables, even if it consists of only one pitch movement (p. 79).

In the GDI, three such typical pitch configurations are distinguished. First there is the *hat pattern*, which was described in Cohen & 't Hart (1967) as follows:

> It is composed of, from left to right, an initial gradual fall-off, to be called declination, a steep rise, an upward shifted segment of declination line, a steep fall, and a final declination which is the extension of the initial declination line (p. 183-184).

Collier (1972) points out that this pattern can have several free variants, depending on the presence or absence of declination and the number of accents to be realized. As such, the most essential characteristic of the hat pattern is a configuration which consists of an early rise 1, followed by a late fall A.

The second typical configuration of pitch movements is known as the *cap pattern*, which is described as:

> On the lower declination line ("0") is superimposed a sudden rise that is situated late in the accentuated syllable (Rise 3), followed by the contribution of the declination line at a high level ("Ø") and a very late fall in the utterance final syllable (Collier, 1972: 87).

Finally, there is the *valley pattern*, the general shape of which is described as follows:

> On the lower declination line is superimposed a rise (Rise 1) immediately followed by a non-final fall (Fall B) and a stretch of inclination (Rise 4) that results in a peak from which the pitch suddenly drops back to the level of the declination line (Fall A) (Collier, 1972: 87).

The perceptual distinctiveness of these three patterns was investigated in a number of perceptual experiments. In the first experiment (Collier, 1972), three sets of four sentences were read by a native speaker of Dutch. Each set of sentences was realized with a different intonation pattern, i.e. the hat, the cap and valley pattern. These utterances were combined into pairs and recorded on tape. The stimuli pairs were such that an intonation pattern was either matched with itself or with one of the other two patterns. This tape was presented to 16 subjects, who were required to judge the intonational resemblance between the patterns in each pair in terms of the labels 'resembling' or 'non-resembling'.

The results of this experiment indicate that these melodical patterns are perceptually distinguished by subjects with a high degree of accuracy. Matching patterns were identified in 89% of the cases, whereas non-matching patterns were recognized as different in 82% of the cases.

The second experiment consisted of a matching task. A set of utterances was recorded on language master cards(2), each utterance having one of the intonation contours mentioned above. These cards were grouped into quartets in such a way that each quartet contained two intonationally resembling contours. One of these contours was used as a reference stimulus. The other contours were different. In the test, 10 subjects were asked to match the reference stimulus of each quartet with one of the remaining contours. The results entirely corroborate those obtained in the first experiment, in that the matching counterpart of the reference stimulus was correctly identified in 83% of the cases.

The third experiment can be considered as an attempt to generalize the above results to conditions where subjects were presented with a larger variety of intonation contours. For this purpose, 20 utterances were selected from a previously recorded corpus. This selection was pseudo-random, in that the only criterion was that each contour should have at least one resembling counterpart. These utterances were

---

(2) Language master cards can be used on taped-card recorders, which provide easy and random access to recorded utterances. Each card has a short stretch of tape attached to it on which utterances of a few seconds in length can be recorded, while an identification of the stimulus can be written on the cards themselves. The utterance is played when the card is inserted in the tape recorder (Catford, 1977).

recorded on language master cards and eight subjects were asked to find the resemblances and group the cards on this basis. A cluster analysis on the results revealed that subjects distinguished between three groups of contours. One groups consisted entirely of cap patterns, while the second group contained hat patterns only. The third group contained a variety of contours which shared a variety of features such as the small excursion size of the pitch movements.

From these experiments, Collier (1972) concludes that the hat, cap and valley contours constitute three perceptually distinct pitch patterns which are indicative of three abstract intonation patterns.

Regarding the precise number of underlying pitch patterns, 't Hart & Collier (1975) suggest that:

> The total number of discernible intonation patterns must be fairly low, possibly not exceeding 10. At any rate, it must be many times less than the number of pitch contours that can occur in any given corpus (p. 254).

As to the underlying mechanisms that operate to reduce the number of observed pitch contours to a set of abstract intonational categories, 't Hart & Collier suggest two principles, the first one of which is the *elasticity* of contours. This literally means that contours can be stretched or compressed: e.g. a pointed hat contour (1&A) belongs to the same abstract pattern as a contour in which the rise and the fall are separated by several syllables (10A). The second refers to the *optionality* of pitch blocks in a contour. A contour which consists of an endblock only is regarded as deriving from the same intonation pattern as one with the same endblock, but preceded by one or more prefix blocks.

It is undoubtedly the case that the principles of elasticity and optionality play a role in the reduction of all the pitch contours that can be observed into a small number of intonational categories. But further research concerning these aspects is required.

1.8. Intonation and Prominence in GDI.

It should be quite clear by now that intonation is regarded as an essentially perceptual phenomenon and consequently is defined in perceptual terms in Collier (1979) as 'the ensemble of perceivable pitch variations in speech' (p. 357). As it stands, this definition differs considerably from various previous definitions, in mainly two respects. From a melodical point of view, every perceivable pitch event belongs to the realm of intonation. This is unlike for instance Bolinger (1958) and Thorsen (1988), who classify pitch into intonation- and accent-related phenomena. From a functional point of view, the definition avoids reference to the function of intonation in the language system, an aspect which has often been used to distinguish adequately between intonation and tone in tone languages.

Although GDI correctly starts from the assumption that accent and intonation are primarily related to the same phonetic correlate of F0 variation, the fact that some F0 variations are related to the linguistic category of prominence whereas others are related to intonation, has to be accounted for. This is achieved by indicating that pitch contours consist of pitch movements, some of which are prominence-lending, whereas others are not. Hence, GDI postulates an intricate relationship between accent and intonation in utterances, since there are no compelling reasons to distinguish the two phenomena on a purely phonetic basis:

> (...) since this impression of prominence is caused, as a rule, by intonational means to the effect that a prominent syllable is accompanied by an audible pitch movement, there is no way to separate stress and intonation on perceptual grounds. Clearly, the distinction between stress and intonation can only be made within a linguistic framework (...) The study of actual speech events, however, cannot overlook the intricate interplay of stress and intonation at each level of the speech communication process (Collier, 1974: 23).

In this aspect, GDI differs considerably from, for instance, contour interaction models of intonation, in which both phenomena are treated separately. The intonation component provides the overall intonation contour, whereas an accent component provides the local F0 specifications.

A further question about the relationship between accent and intonation concerns their precise mode of interaction. IPO researchers argue that a simple additive model is inadequate. Such model regards pitch contours as the simple linear addition of F0-variations, some of which are related to accentuation, whereas others are related to intonation. The predictions of such a model are stated in 't Hart & Collier (1979):

> Those pitch movements that co-occur with prominent syllables are entirely and exclusively related to accentuation, the remaining pitch movements of the contour are associated with intonation (p. 397).

The main argument against such an additive model is that it does not account for the choice of prominence-lending pitch movements in contours. The model can explain why (8) is a well-formed contour in Dutch. However, it cannot account for the ill-formedness of (9):

(8) Grootmoeder gaat met de kinderen naar het zwembad.

(9) Grootmoeder gaat met de kinderen naar het zwembad.

The additive model implies that in (8) the first rise and the last fall are exclusively related to accentuation, since they occur on the prominent syllables. The stretch of high declination between the two pitch movements is related to intonation. In (9), the rise and fall also occur on the syllables of the prominent words. It does not account for the fact that the fall on 'grootmoeder' and the rise on 'zwembad' yield an ill-formed Dutch intonation contour. This observation suggests that the choice of prominence-lending pitch movements is not entirely free, but is determined by certain constraints.

In order to account for this, the suggestion is that the choice of the pitch movements associated with accentuation is made dependent on the kind of intonation contour that is to be realized in the utterance. Thus, intonation is made dominant over accentuation, as stated in the following principle:

> (a) The nature and the order of all the pitch movements in an utterance are determined by the intonation pattern.
>
> (b) Among the pitch movements of any intonation pattern there is at least one which possesses such phonetic properties as are necessary for bringing about a pitch accent.
>
> (c) The location of the accent-lending pitch movement(s) is determined by the position of the words that carry sentence stress, and more specifically, by the position of the lexically accented syllable in each of these words.
> ('t Hart & Collier, 1979: 400).

This principle abandons the idea that pitch movements can be entirely and exclusively related to accentuation and it accounts for the correctness of the choice of pitch movements to signal prominence. This principle predicts that if the same intonation pattern is realized and the accentual requirements change, the only difference will relate to the position of the prominence-lending pitch movements, not to their nature. The intonation pattern that is realized in (10) to (13) is one in which an accent-lending rise is followed by an accent-lending fall. If the location of the pitch accents change, this basic intonational structure is maintained:

(10) Grootmoeder gaat met de kinderen naar het zwembad.

(11) Grootmoeder gaat met de kinderen naar het zwembad.

(12) Grootmoeder gaat met de kinderen naar het zwembad.

(13) Grootmoeder gaat met de kinderen naar het zwembad.

In all these examples, the basic intonational structure (rise followed by a fall) is respected. Only the location of the rise and fall change according to which words in the utterance are made prominent.

It should be noted that this principle also accounts for the ill-formedness of (9), since in this case the basic intonational structure is violated: the fall precedes the rise. The primacy of intonation over accentuation hence makes it possible to account for the type and the location of the prominence-lending pitch movements in a very elegant manner.

2. Target Analyses of Dutch Intonation.

In this section, two target analyses of Dutch intonation are summarized, which have originated as reinterpretations of the intonational data provided by GDI. Contrary to GDI, which is a phonetically motivated analysis, these analyses are essentially phonologically inspired: they attempt to capture phonological generalizations underlying the phonetic data. In doing so, they abstract away from the concrete acoustic details of F0 contours, without getting out of touch with the phonetic data.

2.1. The Peak-Feature Model.

2.1.1. General characteristics.

The preliminaries of a peak-feature model of intonation were laid out in Ladd (1983a, 1983b). This model attempts to reconcile the goals of linguistic generalization and phonetic description in intonation research. The feature model -like GDI- essentially adopts a tone sequence approach to intonation, in the sense that intonation contours are analysed as a sequence of structurally significant pitch events. Ladd's analysis however differs fundamentally from the GDI, since these pitch events are specified as points (or targets) rather than pitch movements. Ladd argues that only targets are relevant for describing the structure of intonation. The pitch movements that can be observed in F0 curves do not matter in a phonological perspective: they are the automatic result of transitions between the linguistically relevant targets:

> (..) syllable contours must be seen as the incidental result of transitions and targets: what counts phonologically is the alignment of a syllable, and the associated contour peak - not the shape of the contour chunk that accompanies the accented syllable as the pitch moves from one point to another (Ladd, 1983b: 730).

The structurally significant points that are distinguished in the model are associated with accent peaks, accent valleys and syntactic boundaries.

Another important aspect of the model is that the targets are considered as the location of a tone: i.e. 'a phonological (intonational) segment characterized by a bundle of features' (Ladd, 1983b: 728). The tones can occur individually as either HIGH (H) or LOW (L), but can also occur in fixed combinations as higher level units, such as HL or LH. All tones are further specified as a bundle of features and are interpolated by F0 transitions.

As to the number of features required for characterizing tones, Ladd starts from the assumption that there are in fact only two dimensions along which a tone has to be specified: F0 and time. Since tones specify points rather than contours, it is sufficient to specify a tone in terms of its peak height and its alignment(3) in time to the

---

(3) The feature model distinguishes between 'association' and 'alignment'. 'Association' refers to the relationship at the abstract phonological level between the tone and the accented syllable. 'Alignment' refers to the positioning in time of the tone, which is determined by a number of factors, such as the specification of phonological features.

segmental string. For peak height, there is a basic phonological opposition between H and L (which are not the tones themselves, but have to be considered as features of tones). Furthermore, a peak may be raised, for which the feature [RAISED PEAK] is introduced, or downstepped captured by the feature [DOWNSTEP]. Peak alignment on the other hand, can be expressed by means of a feature [DELAYED PEAK].

In the next paragraph, we summarize Ladd's argument for how these features can be used successfully to account for a number of intonational phenomena in languages and to make meaningful statements of cross-classification.

2.1.2. Targets and Features.

The fact that Ladd specifies contours in terms of targets rather than pitch movements derives, for instance, from word-accent phenomena in Swedish. The dimension of *peak alignment* can be shown to be relevant in this respect. It is observed that word accents 1 and 2 are realized in phonetically different ways across Swedish dialects. An accent can occur as a rise in one dialect, whereas it is a fall in another. In a pitch movement approach, the two configurations would have to be treated as different. A target analysis can capture this variation by postulating a basic underlying HL accent. Cross dialect differences can then be expressed in terms of peak alignment: when the H-target occurs late in the syllable, a rise is obtained; if the H is aligned early in the syllable, a fall occurs across the syllable. Hence, it can be shown that two phonetically different pitch configurations are in fact the realization of the same underlying phonological unit. It is evident that a feature [DELAYED PEAK] captures this phenomenon very well from a phonological point of view. At the same time, it does credit to the phonetic details.

A similar treatment of scooped contours in English and German provides the same general insight. Both English and German have contours which function as falls, but which begin with a rise on the accented syllable: they reach their peak only in the following unaccented syllable. A feature treatment of this phenomenon is able to explain why both contour types have related functions and describe their phonetic shape. This is achieved by postulating that all falls are HL, with [-DELAYED PEAK] for ordinary falls and [+DELAYED PEAK] for scooped falls. The basic similarity between the two contours is expressed, while the rising part of the scooped fall is distinguished from a rise in LH.

As to the dimension of *peak height*, we mentioned the basic contrast between H and L. Furthermore, two features can be shown to be relevant, i.e. [DOWNSTEP] and

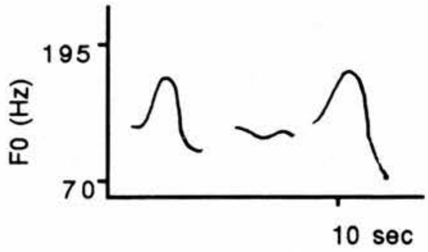
[RAISED PEAK]. Downstep, originally applied to the analysis of intonation by Pierrehumbert (1980), applies to contours that are 'characterized by a rather steep overall downward slope through the individual accent peaks' (Ladd, 1983b: 733). The accents are downstepped, as in a number of African tone languages. The downstepped contour can be characterized by selecting a separate feature [DOWNSTEP]. This makes it possible to formulate an explicit relationship between the downstepped contour and its non-downstepped counterpart, as in (14) and (15):

(14) I really don't think he ought to be doing that
H HL L%

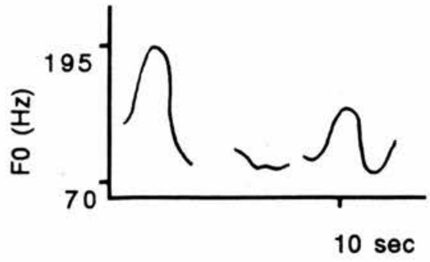
(15) I really don't think he ought to be doing that
H !H !H !HL L%

[RAISED PEAK] can account for the height of the last peak in (16) as compared to (17):

(16)



(17)



Ladd argues that the effect of [RAISED PEAK] is to be distinguished from the effects of the use of pitch range, which also raises F0. It is for instance the case that in a wider pitch range, all the peaks of the utterance are raised. But in both the normal and widened pitch range, a [RAISED PEAK] is scaled higher than the others in the same way.

2.1.3. A Feature Analysis of Dutch Intonation.

In his analysis, Ladd postulates a distinction between two basic types of pitch accents: a H accent, corresponding to the Type 1 rise in the GDI, and a HL accent that can be regarded as the equivalent of the Type A fall. The specification of the relevant distinctive features for these accents, ultimately yields all the relevant pitch movements of the GDI (table 2.3).

IPO	Feature Analysis	
	Symbol	Explanation
A	HL	Pitch accent
B	%	Prosodic boundary, followed by a return to normal pitch of phrase initial unaccented syllables
C	L%	Low boundary tone
D	-	Transition from H to L
E	!H	[+downstep]
1	H	Pitch accent
2	H%	High boundary tone
3	H	[+delayed peak]
4	-	transition from L to H
5	HL	[+raised peak]

Table 2.3: Summary of the reinterpretation of pitch movements in the GDI into features.

At first sight, this seems just a reformulation of the GDI terminology. But Ladd shows that such a feature analysis of Dutch intonation is well equipped to express linguistic generalizations. This can be illustrated with respect to the flat hat pattern, which he argues, underlies a great number of Dutch pitch contours. The relationship between these contours can be established by postulating a H_0__HL pattern, i.e. a HL accent, optionally preceded by any number of H accents, as in (18) to (20):

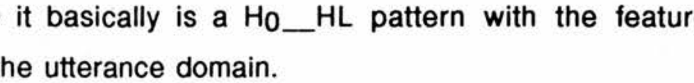
(18) Dat huis is te koop
H HL L%

(19) Dat huis is te koop
HL L%

(20) De pianostemmer is vanmorgen geweest
H % H HL L%

Although the GDI analysis of (18) as 10A, (19) as 1&A and (20) as 30B010A effectively models phonetic reality, it is unable to express the similarity between these contours, which are intuitively regarded as related by native speakers. The feature model however formalizes their similar identity: a nuclear HL, optionally preceded by one or more H's. This is a pattern which can be interrupted by a syntactic boundary as in (20).

The feature [DOWNSTEP] can account for contours as (21):



(21) Al die ingewikkelde regelingen zijn afgeschaft.
H !H !H !HL L%

This enables the model to show its basic relationship with other declarative patterns, since it basically is a H_0__HL pattern with the feature downstep added throughout the utterance domain.

The feature [DELAYED PEAK] is useful in the analysis of contours containing a rise of Type 3, as in (22):

(22) Ben je helemaal niet gegaan
H L%
[+DELAYED]

The utterance final C-movement is interpreted as a low boundary tone.

Finally, the feature [RAISED PEAK] accounts for the occurrence of a smallish pitch movement type 5, preceding a final fall in utterances as (23):

(23) Jan heeft een nieuwe auto
H HL
[+RAISED]

The gradual pitch movements D and 4 which are regarded as relevant pitch movements in GDI, are analysed in the feature model as gradual transitions between H and L or vice versa.

It should be clear that a feature analysis of Dutch intonation is a powerful formalism to express linguistic generalizations: pitch contours which are phonetically quite different, can be shown to be the realization of the same underlying pattern, which can be correlated to a particular function, such as 'neutral declarative'. Moreover, pitch patterns that have a different function can be shown to be the realization of a distinctly different phonological pattern.

2.2. An Autosegmental Model.

2.2.1. General Characteristics.

A more comprehensive target analysis of Dutch intonation was undertaken by Gussenhoven (1988a, 1988b). He shows that a model, which was originally developed for the description of British English intonation in Gussenhoven (1984), can be successfully applied to Dutch intonation. It distinguishes between two tonal elements (H and L) which combine into three basic tonal morphemes (tones): H^*L, L^*H and H^*LH(4) . One of these morphemes has to be chosen at each accented position

---

(4) The asterisk refers to the accent. In Gussenhoven's notation, it appears above the accented tonal element. However, typographical constraints do not permit this

in the segmental string, in such a way that there is a one-to-one correspondence between accents on the tonal tier and accented positions on the segmental tier. The large variety of pitch contours in Dutch is derived by phonological rules. In some cases however, the tones surface in a more straightforward way.
Association between the segmental and suprasegmental tier is established conform to Goldsmith's Well Formedness Condition:

(24) a. Association is exhaustive: i.e. no segments on either tier are left unassociated.
b. Association lines do not cross.

Gussenhoven remarks that a rigorous application of this version of the Condition would lead to a number of difficulties. The first difficulty is related to the fact that association is to be exhaustive. If the model is to be able to generate pitch configurations of unaccented stretches before an accented syllable, it should be assumed that a tone T^* cannot spread leftwards. Such a solution would avoid a situation in which H^*L and H^*LH are always preceded by a H and L*H always by a L tone. This solution is problematic, since segmental material to the left of the first accent is then unassociated, because there are no tone elements to the left of the accented element in each tone. In order to avoid this problem, Gussenhoven modifies the principle of exhaustive association:

(25) Association is exhaustive, but only from * onwards to the domain end.

The second problem results from the fact that association of tones is not governed by tone-bearing units as in tone languages: i.e. tonal elements are associated with points in the time domain which cannot be equated with phonological units in the segmental string. It is suggested by GDI, that pitch movements can be classified with respect to their timing with respect to the syllable. Moreover, there are no restrictions on the number of tonal elements that associate with a syllable, unlike in some tone languages, which can also be regarded as indicative of the absence of tone-bearing units. Hence, conventional association is impossible, because reference can be made to segments in the tonal tier, but not to segments in the segmental tier. Therefore, Gussenhoven introduces a modification, which allows for a correct association of the tones to the segmental string:

---

convention and, hence, the asterisk has been placed immediately on the right of the accented tonal element.

(25') Only the second tone element of a tone can spread.

This principle is illustrated in (26):

(26)

Besides these basic tonal elements and tones, the model incorporates a small number of *phonological rules* and *modifications*, the difference between the two being the fact that rules contribute to the formation of larger intonational domains (tone units), whereas modifications operate on single tones.

2.2.2. Phonological Rules.

2.2.2.1. The Tone Linking Rule.

The application of the Tone Linking Rule (TLR) is formulated as follows:

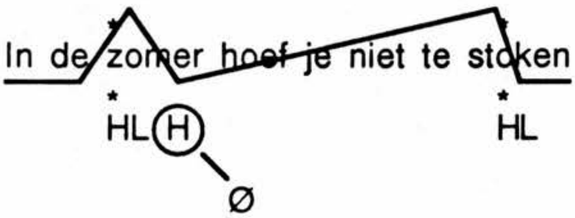
> The final tone element of the left-hand tone, together with its domain-end boundary, is deleted, causing the two tone elements that are now adjacent to be connected directly (Gussenhoven, 1988a: 236).

Formalized: TLR: $T \rightarrow \emptyset / \underline{\qquad} T^{*}$.

This is exemplified in (27) and (28):

(27) a.

b.



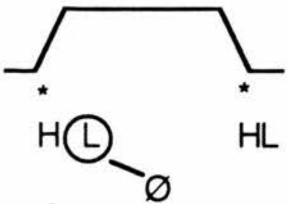
(28) a.

b.

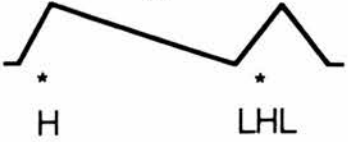
In both cases the (a) and (b) contours are related to each other in the sense that the (b) contours are said more quickly than the (a) contours.

It should be mentioned that it is possible to have *partial linking*, in which case the deleteable tone element is shifted to the right rather than deleted completely, as in (29):

(29) Complete linking :



Partial linking :



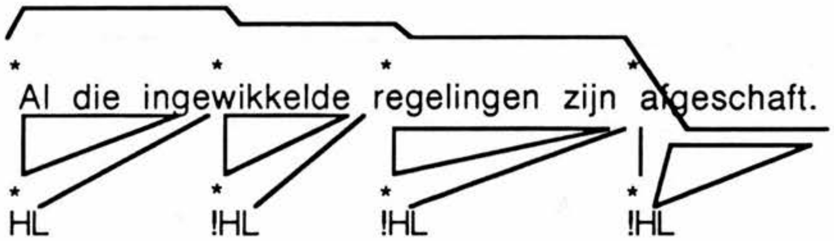
2.2.2.2. Downstep.

The Downstep Rule is a phonological operation, which applies in a particular environment: it is optionally activated in sequences of H^*L tones, thus creating a terrace-shaped contour, with the steps at the positions of the accents. It basically consists of two separate operations. In the first one, the H^* following a H^*L is downstepped. Secondly, every H^* except the final one spreads:

DOWNSTEP: a. $\varnothing \rightarrow ! / H^*L$ ____ H^*
b. Spread $H^* /$ ____ LH^*

This is illustrated in (30):

(30)



The L after the spread H^* is pushed on the following H^*, where its effect is masked by the step down to $!H^*$.

The formulation of this rule is clearly at odds with association principle (25'), which states that only the second tone element can spread. Therefore, Gussenhoven reformulates the principle in such terms that only the last tone element of a tone can be a boundary tone, which implies that spreading can only apply to boundary tones or tones which only force the boundary tone to the domain end.

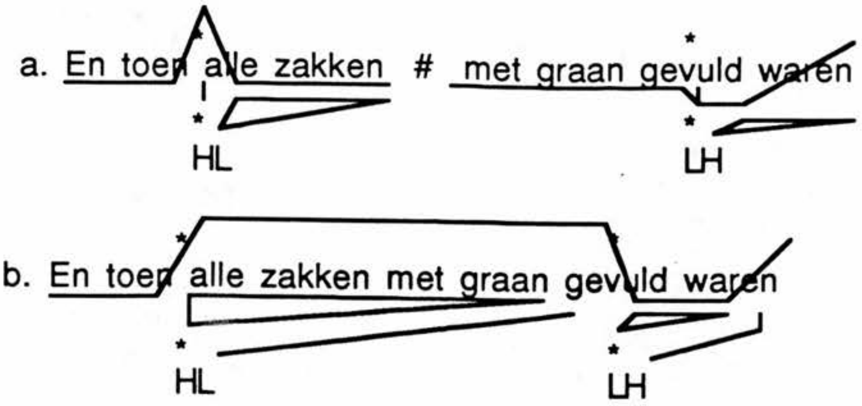
2.2.2.3. Narration.

This rule is essentially stylistic in nature and its application seems to be confined to reading aloud. It can be stated as follows:

NARRATION: Spread T^*.

Its application is illustrated in (31):

(31)



Narration cannot be applied to a $H^{*}LH$ sequence. It also implies the deletion of the domain boundary between the adjacent tones involved in an operation, which is also the case in Tone Linking and Downstep. In Narration moreover, the tone element after T^{*} is displaced to the right as a result of the spreading of T^{*}.

2.2.3. Modifications.

Each tonal element of the tonal morpheme can be optionally affixed with a *modification*. Gussenhoven (1988a) points out that the phonetic effect of affixation is fully integrative i.e. "the surface manifestation of a modified tone cannot be divided into a segment that represents a tone and a segment that represents a modification" (p. 239). The model distinguishes between three modifications.

2.2.3.1. Half-completion.

In half-completion, the second tone element is changed to a mid-level tone (M), formalized as:

HALF-COMPLETION: $T \rightarrow M / T^{*}$ ___

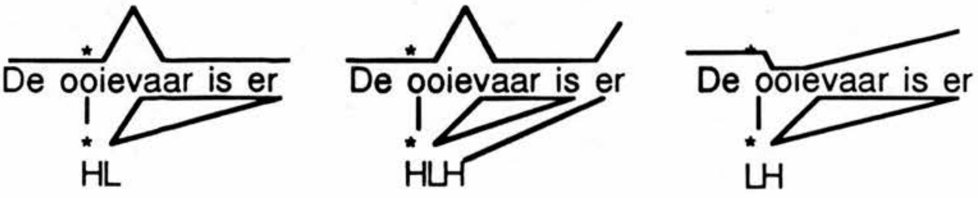
This is illustrated in (32):

(32)

2.2.3.2. Delay.

This modification shifts the association of a target of a starred tone element to the right. Delay should not be taken to mean that the accent in the segmental string is shifted. Furthermore, it is also distinct from spreading, because there is only one single timing point with which the tone associates (33):

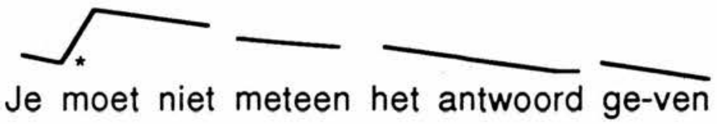
(33)



2.2.3.3. Stylization.

The final modification is stylization. This modification is generally characterized by a lengthening of all the syllables in the domain of a tone, starting from the accented one. Furthermore, there tends to be plateau formation, as in (34):

(34)



3. General Discussion.

In this chapter, three approaches to the analysis of Dutch intonation have been summarized. First, the Grammar of Dutch Intonation, developed at the Institute for Perception Research, was discussed. This model results from a typically form-driven, bottom-up approach to the study of intonation, i.e. it takes its starting point in a large set of F0 observations and aims to describe these in an explicit manner. In this process, it leaves any functional aspects out of consideration from the start, i.e. the acoustic parameters that are observed in the speech signal are regarded as indications of more abstract formal categories rather than of concrete functional ones. Consequently, it is an explicit aim of the GDI to abstract away from the concrete level of analysis to model the underlying formal system.

A first degree of abstraction is obtained by applying the stylization method to a large number of utterances. In this method, F0 curves are modelled artificially by a minimal set of straight lines in such a way that the artificial contours are perceptually indistinguishable from their corresponding originals. Such artificial contours are assumed to contain information related to intonation only. Application of this method has led to the discovery that global contours at an atomistic level consist of a small number of discrete pitch events: these are known as perceptually relevant pitch movements.

A second degree of abstraction is accomplished by standardizing the relevant pitch movements with respect to the parameters of their excursion size, their slope, their location in the syllable and their duration. This was achieved by experiments into the tolerance levels of native speakers with respect to each of these perceptual dimensions. Contours which are made up of such standardized movements are not necessarily indistinguishable from their corresponding originals, but can be regarded as perceptually equivalent.

The next stage consists of accounting for the combinatory possibilities of pitch movements in pitch contours. This makes an intermediate level of description necessary, i.e. the pitch block. Different types of pitch blocks are distinguished depending on their typical location in a global contour. As such, it is possible to formulate an explicit hierarchical relationship between the global and atomistic aspects of pitch contours: global contours are accounted for in terms of a combination of pitch blocks, which in turn are defined at the atomistic level by their constituting pitch movements.

At the third and highest level of abstraction, the pitch contours are assumed to be related to a small number of abstract intonation patterns. The nature and number of these patterns has not been investigated exhaustively. Nevertheless, it is postulated that the relationship between these patterns and their concrete realizations operates on the principles of optionality and elasticity.

The relationship between all these aspects in the GDI at the concrete/abstract and atomistic/global level is schematized in figure 2.4:

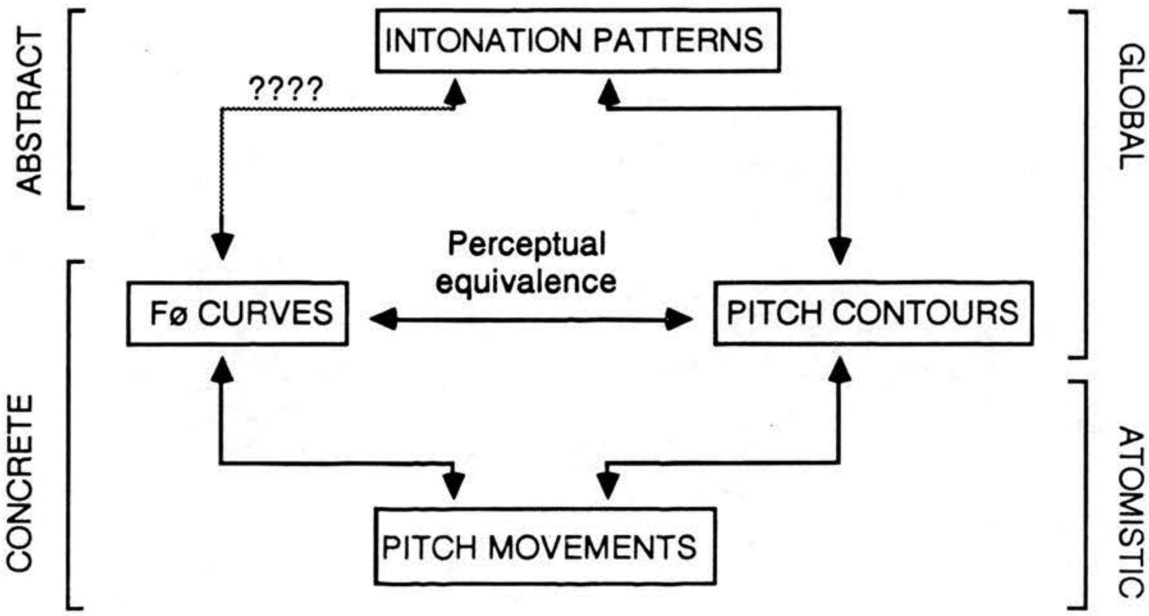


Figure 2.4: Schematic representation of the relationship between the concrete/abstract and the global/atomistic aspects of intonation. The question marks indicate doubt about whether a direct link between F0 curves and abstract intonation patterns can be established (adapted from 't Hart & Collier, 1975).

While the GDI can be characterized as an abstract phonetic model of intonational form, the peak-feature model (Ladd) and the autosegmental model (Gussenhoven) are primarily phonologically motivated. Both models rely on the phonetic data provided in the GDI and account for their linguistic structure in terms of tonal targets. In the peak-feature model, these tonal targets are specified as bundles of features, which determine the concrete phonetic realization of the F0 peaks. The relevant features are [DELAYED PEAK], [RAISED PEAK] and [DOWNSTEP]. In the autosegmental model, tonal targets are associated with accented syllables and an appropriate underlying structure is derived by a small set of phonological rules (tone linking, downstep, narration) and tonal modifications (half-completion, delay and stylization). In both models, the concrete phonetic form derives from a simple linear interpolation between the tonal targets.

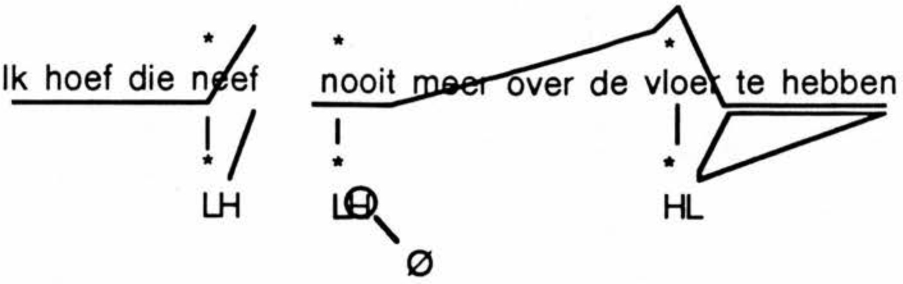
The models that have been discussed clearly differ in their theoretical pretentions: GDI attempts to account for the observed intonation patterns in *abstract phonetic* terms. The peak-feature and autosegmental models on the other hand aim to identify the underlying structural units of contours and account for the data in terms of these units. Consequently, they can be shown to have greater explanatory power and

are better equipped to characterize intonational contrasts. Both aspects will be discussed briefly on the basis of examples taken from Gussenhoven (1988a).
As to its explanatory power, the autosegmental model is capable of explaining certain prosodic phenomena, whereas GDI can only list these phenomena as relevant and is essentially unable to provide an explanation for them. Two of Gussenhoven's examples clearly illustrate this. The first relates to a succession of two accents: L^*H followed by $H^*L(H)$ without application of the TLR. In this instance, the pitch can be observed to drop at the end of the domain of L^*H, illustrated in (35):

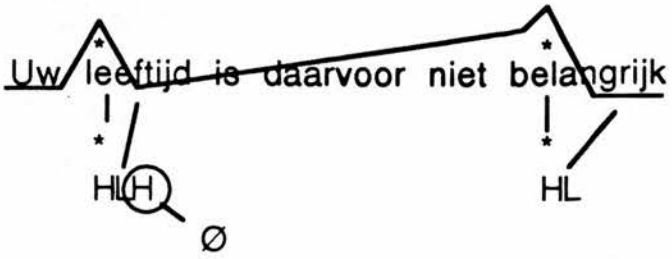
(35)

This pitch reset follows necessarily from the underlying phonological representation and consequently does not have to be specified separately. A similar treatment of the phenomenon is found in Ladd's feature model. GDI conversely, regards the pitch reset at the prosodic boundary as a relevant pitch movement 'B', since it is manifested as such in the surface structure. Moreover, the autosegmental model explains why 'B' is not accent-marking (it is not associated with a starred tone) and why it occurs at a syntactic boundary (association is exhaustive from the starred tone onwards to the domain end, which coincides with a syntactic boundary). GDI can only observe these characteristics and is not able to account for them formally.
The second example relates to the status of the GDI movement '4', which in some cases marks an accent, whereas in other circumstances it does not do so (36) - (37):

(36)



(37)

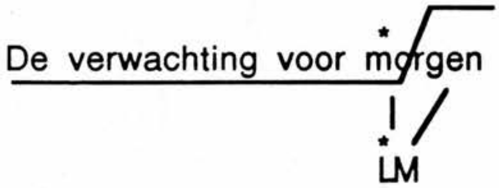


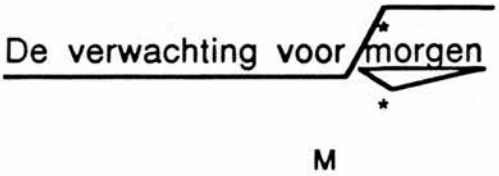
The autosegmental analysis is able to account for the different behaviour of '4' in both examples, because it arises from different underlying representations. In a manner similar to Ladd's analysis, '4' is not regarded as a separate phonological unit, but as a transition between two tonal segments, which results from the application of the TLR. In (36), '4' marks an accent on 'nooit', because it links an underlying L^* tone with a H^*L. In (37), '4' links an unstarred L tone with H^*L. Gussenhoven shows that when it comes to characterizing intonational contrasts:

> the phonological analysis quite naturally characterizes the phonological difference between a number of phonetically different contours for which GDI has only got one description available (Gussenhoven, 1988a: 248).

In order to argue his case, Gussenhoven adopts the technique of close-copy stylization, in which for a given GDI contour, he creates a phonetically slightly different contour. These minimal pairs were evaluated auditorily. Gussenhoven concludes that "the two members of the pair were nowhere near being 'perceptually equivalent' and appeared to represent categorically different contours" (Gussenhoven, 1988a: 248).

A contour described as 1 can be analyzed phonologically in two different ways. It can represent either a half-completed L^*H or a stylized LH^*. For a half-completed L^*H, the rise represents the movement from L to M. Consequently, the rise will take place within the accented syllable. For a stylized LH^*, the rise represents the movement from the baseline to the mid-level tone, which is associated with the accented syllable. Therefore, the rise occurs before the accented syllable.

(38)


(39)


Although the perceptual effect of both operations is said to be very distinct and seems to correlate with a semantic difference, GDI cannot analyse the rise as other than '1'.

The explanatory power and the success in characterizing intonational contrast of both the peak-feature and autosegmental model is largely due to the assumed nature of the tonal segments. While the GDI uses pitch movements as its basic descriptive units in the analysis of Dutch intonation, the phonological models postulate the existence of tonal targets. Phonetically, a pitch movement analysis is almost self-evident in that pitch contours manifest themselves as a combination of F0 transitions in the acoustic signal. Gussenhoven (1988b) argues that the relevance of tonal targets can be shown phonologically by the fact that it is possible to formulate well-motivated rules of intonation which refer to these assumed targets and agrees with Liberman that:

> (...) we might argue that a particular theory based on the hypothesis of static tonal segments is able to predict a wide range of observed intonational outputs on the basis of simple and consistent underlying representations acted on by a set of well-motivated rules (Liberman, quoted in Gussenhoven 1988b: 317).

Concrete phonetic evidence for the relevance of such tonal targets is nevertheless very difficult to provide. Chapter 1 (section 2.4) demonstrates that, physiologically, speakers need a minimal amount of time to produce melodic variations (Sundberg, 1979). It should be observed here that in speech, the rate of melodic changes is actually considerably slower than the maximally attainable speed. This can arguably be taken as evidence that speakers do not program their melodic changes as pitch jumps between pitch levels (Collier, 1983). Hence, at the level of speech production, an analysis of intonation in terms of pitch movements does justice to this observation.

Also the perceptual reality of targets has not been shown conclusively. In Verhoeven (1987)(5), a discrimination experiment is reported which aims at establishing the sensitivity of informants to variations in tonal targets assumed to underlie particular pitch movements. A hypothesis is elaborated that the perceptual sensitivity of informants to variations in F0 onsets and offsets of pitch movements depends on the nature of the phonological segment associated with the movement and on which part of the movement is manipulated. It is predicted that if an accented syllable is associated with a H tone realized as a rising movement, small variations in the F0 target should be clearly perceivable, whereas variations in the F0-onset of the movement towards this target should not really have any perceptual consequences.

The onset frequency and the offset frequency of rises were manipulated to test this prediction. The rises of the stimuli were located early in utterance-initial syllables and constituted part of the hat pattern. The stimuli from each phonetic continuum were paired into AX discrimination items, the difference between the values of the variables being 3 steps. That is to say that in the onset condition, there was a 4.5 semi-tone difference between the onset frequencies of the movements in stimuli A and X. In the endpoint condition, the difference between the endpoint frequencies of the rises in A and X amounted to 2.5 semi-tones. 20 native speakers of Dutch were presented with the stimulus pairs from both continua and were asked to judge whether they could hear a difference between them or not.

The results show that there is a significant difference in discrimination scores between both conditions. In the onset condition, only 19% of the informants report hearing a difference. In the endpoint condition on the other hand, 62% of the informants report to hear a difference, even though the physical difference in these stimuli was only half as big as this in the stimuli of onset condition. This differential sensitivity of informants to variations in onset and offset of rises is taken as evidence for the reality of tonal targets in intonation.

A second relevant experiment is reported in House (1985, 1987). He investigated the categorization strategies of informants in the perception of pitch complexes. His aims are two-fold. On the one hand, he attempts to provide an insight into whether informants categorize pitch complexes on the basis of a continuous pattern storage

---

(5) The complete paper can be found in the Postscript section.

technique in terms of pitch movements or whether categorization is achieved by the storage of discrete pitch frequencies. On the other hand, he investigates the possible influence of spectral changes on these categorization strategies.

For the purpose of this experiment, a Swedish /a/ vowel was synthesized, in which F0 was manipulated to create 18 stimuli, that represented three pseudo-continua. In all instances, the F0 onset was kept constant at 120 Hz, whereas the F0 endpoints in the three continua were different at 100, 160 and 120 Hz respectively. The stimuli in each continuum contained an F0 turning point in the middle of the vowel, and the height of the turning point was varied between 80 and 180 Hz in steps of 20 Hz. The characteristics of these stimuli are illustrated in figure 2.5:

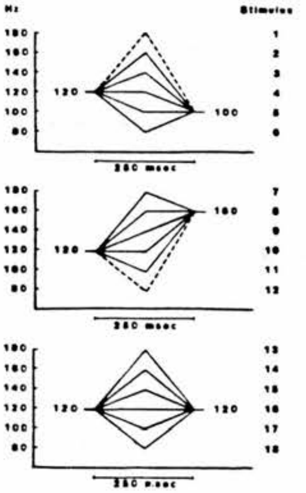


Figure 2.5: Stylized contours for the stimuli with a steady-state vowel. The dashed lines (i.e. stimuli 1 and 12) represent the contours of stimuli A and B of the ABX configuration (Adapted from House, 1985).

The stimuli were combined into ABX items, in which A consisted of a rise-fall complex (stimulus 1) and B of a fall-rise complex (stimulus 12). The X stimulus could be any other stimulus from the continuum.

In order to investigate the influence of rapid spectral changes on the categorization of F0 patterns, three other versions of the test were produced by introducing an intensity drop, preceded and followed by formant transitions, at three different places in the vowel: early, middle and late. The structure of these stimuli is illustrated in figure 2.6:

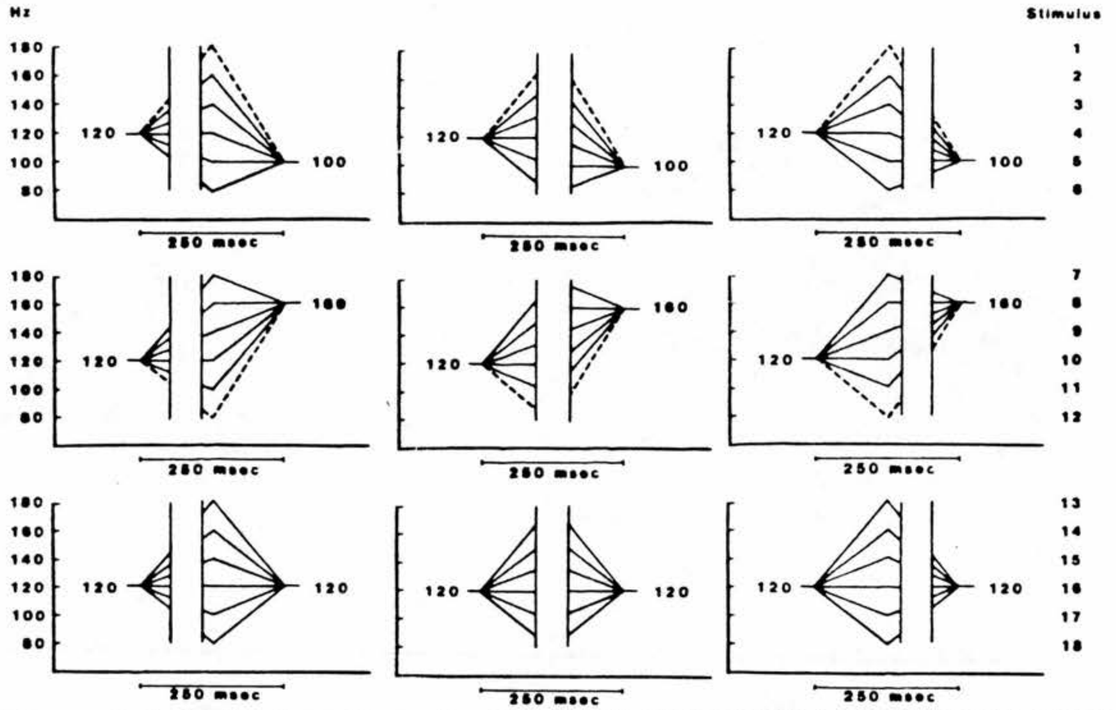


Figure 2.6: Stylized contours for the stimuli with spectral and intensity manipulations. The dashed lines (stimuli 1 and 12) represent the contours of stimuli A and B of the ABX configuration. In (a), the spectral and intensity modifications occur early, in (b) in the middle and in (c) late in the vowel. (Adapted from House, 1985).

The test was administered to 36 native speakers of Swedish, who were asked to listen to the first two sounds of each item and to decide whether the third sound was more like the first or the second sound.

The results of this experiment suggest that all subjects categorize the stimuli on the basis of pitch movement pattern recognition, provided that the vowel is not interrupted by rapid spectral changes. In the presence of spectral modifications, both categorization strategies are used. It is however not clear whether there is any preference for either of these strategies. House concludes that spectral changes clearly affect the perception of pitch contours, in that during spectral changes the perceptual mechanism seems to be less sensitive to pitch movements than during long steady-state portions. He hypothesizes that spectral changes (stops etc.) may act as perceptual boundary markers which enhance the perception of discrete pitch frequencies, while inhibiting the perception of continuous pitch movements.

As far as the general matter of pitch movements vs. targets is concerned, it can be concluded that the concrete evidence for the phonetic relevance of tonal targets is available, but rather limited. Nevertheless, it has to be mentioned that both types of

analysis are not necessarily incompatible in that they just represent different levels of abstraction. A phonetic pitch movement analysis does justice to phonetic reality, whereas from a phonological point of view, the more abstract target analysis allows for interesting generalizations to be made about intonational structure.
Concerning the relationship between intonation and prominence, the models that have been discussed share the basic assumption of tone sequence theories that intonation contours are made up of a sequence of pitch events associated with accented syllables. Unlike contour interaction theories, which regard sentence intonation as a phonological primitive in itself, GDI does not distinguish between pitch events associated with accents and those related to intonation, since "there is no way to separate stress and intonation on perceptual grounds" (Collier 1974: 23). Prominence is thus essentially realized by intonational means.
The peak-feature and autosegmental model specify intonational structure in terms of tonal targets associated with accents and intonational form derives from the linear interpolation between these targets. The latter assumption is rejected in the GDI in that it postulates a hierarchical relationship between abstract intonation patterns and the concrete phonetic realization of pitch accents, which is determined by the general requirements of the abstract intonation pattern. To put it more concretely: in the peak-feature and autosegmental model, the intonation contour derives from the specification of a set of underlying accents, whereas GDI assumes that the specific identity of accents depends on the abstract intonation pattern that is to be realized.

4. Conclusion.

The models of Dutch intonation that have been discussed in this chapter basically approach the analysis of Dutch intonation in essentially different ways. The GDI can be regarded as the most comprehensive phonetic analysis of Dutch intonation to date, in that it has provided an explicit account of the possible pitch contours in Dutch. The experimental techniques developed for this purpose have provided a wealth of attested intonational data so that the conclusions in the GDI are experimentally verifiable. This can be regarded as a major improvement over the traditional impressionistic approaches to intonation analysis. Besides being a model of Dutch intonation, the GDI has also provided a number of essential insights into the perception of intonational features generally.

The peak-feature and autosegmental models are abstract phonological models of Dutch intonation, in which intonational structure is accounted for in terms of tonal targets associated with accented syllables. This approach enables powerful generalizations to be made about intonational form.

CHAPTER 3

Pitch Movement Alignment as a Perceptual Dimension of Intonation.

0. Introduction.

Stated in simple terms, pitch movement alignment refers to the location of the movement with respect to the segmental string. This dimension has been discussed with respect to different languages. In some languages, lexical items may contrast in their 'accent'. In Swedish for instance, some segmentally identical words may be realized with either accent 1 or accent 2, expressing different meanings. It has already been pointed out in chapter 2 that the main phonetic difference between these realizations is related to the timing of the F0 movements which are otherwise identical (Bruce, 1977). This feature has also been found useful to account for characteristic intonational differences between Swedish dialects.

In languages which do not use such an accentual system, there are also pitch movements or pitch complexes "which are categorically different in their location with respect to the syllable boundaries" (Collier, 1983: 244). This is arguably the case for Dutch, for which GDI distinguishes between rises which are located early, late or very late in the syllable. A similar categorization is proposed for falls, although the precise timing differences differ from those of rises. For English, Pierrehumbert & Steele (1987) suggest a two-way categorization of rise-fall complexes, which is also argued to be relevant for German intonation (Kohler, 1987). In this chapter, all these aspects are examined in detail. Subsequently, the relevance of pitch movement alignment to Dutch intonation is discussed.

1. General Aspects of Pitch Movement Alignment.

1.1. Evidence from Speech Production.

The alignment of pitch movements has recently been fairly thoroughly investigated with respect to English (Pierrehumbert & Steele, 1987, 1989, Silverman & Pierrehumbert, in press). This experimental work has mainly concentrated on alignment differences in speech production. Pierrehumbert & Steele (1987) suggest that there are different nuclear rise-fall intonation patterns in English, which differ from each other in the alignment of the F0 peak with respect to an accented syllable. They postulate that this difference reflects a binary linguistic distinction. In one category, a low tone is assumed to be associated with the accented syllable, so that the F0 peak is expected to occur rather late in the syllable. In the second category, the accent is realized by means of a high tone, which is reflected in an early alignment of the F0 peak. An alignment difference of this kind is illustrated in figure 3.1:

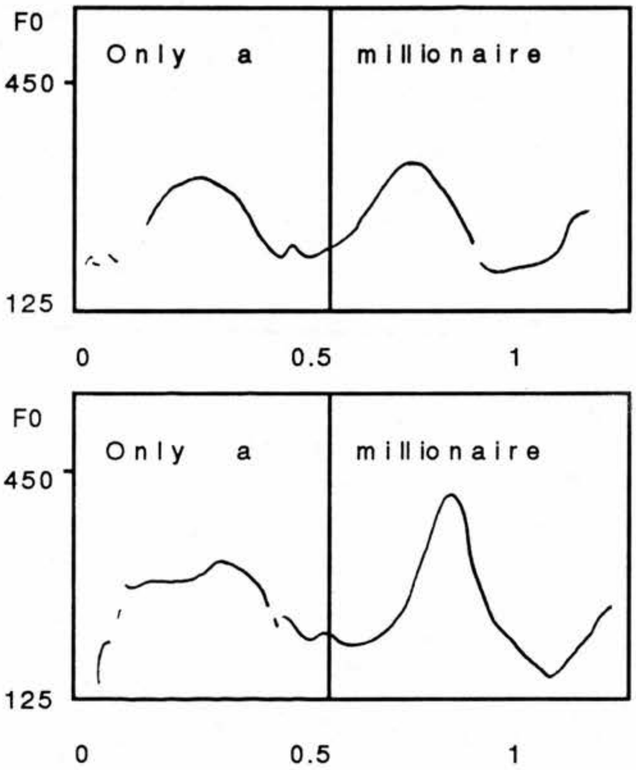


Figure 3.1: Illustration of the F0 peak alignment difference in the utterance 'only a millionaire'. In (a) a high tone is associated with the accented syllable, whereas (b) has a low tone. The peaks clearly occur at different locations in the syllable (Adapted from Pierrehumbert & Steele, 1987).

While Pierrehumbert & Steele assert that this distinction is of a categorical nature, Gussenhoven (1984) proposes a continuous dimension of peak delay with an ideal preferred position:

> (...) while there may be an 'ideal' target for delay (say, one syllable after the nuclear tone if there is one), delay is a gradual modification (Gussenhoven, 1984: 218).

This peak delay feature allows for a wide range of possible locations. It is precisely this question of category vs. continuum that was investigated experimentally in Pierrehumbert & Steele (1987).

For this purpose, a physical continuum of peak delay was constructed on an accented syllable of a short utterance and subjects were asked to imitate each of these intonation patterns. The rationale behind this task was as follows. If the alignment continuum is perceived as continuous by native speakers, the imitations should reflect this continuity in that subjects should be able to imitate each of the individual alignment locations fairly accurately. If however the experimental continuum is perceived in terms of two categories, the imitations should cluster into two distinct groups.

The stimuli for the experiment consisted of the utterance 'only a millionaire', in which the location of the F0 peak of a rise-fall intonation pattern was shifted through the accented syllable 'mil' from an early to a late position, i.e. from 35 to 315 msec from the end of the prevocalic [m]. The peak shift was 20 msec, so that 14 stimuli were obtained. These stimuli were randomized and presented to 5 native speakers of American English, who were instructed to imitate the utterances, paying particular attention to the intonation pattern. These imitations were recorded and F0 was analyzed. Finally, peak delay was measured as the difference between the time of the F0 peak and the time of the [m] release.

Pierrehumbert & Steele report to have found clear bimodal distributions of the peak delay data histograms in three out of five subjects. In one of these subjects, peak delay in the first 9 stimuli varied between 100 and 150 msec, whereas in the last 4 stimuli peak delay clustered between 200 and 250 msec. For the remaining stimuli, the peaks were located somewhere in between these values. The boundaries in the data for the other two subjects were reported to be different. The data for the remaining two subjects showed a rather weak bimodal and a fully unimodal distribution. From these data, Pierrehumbert & Steele (1987) conclude that:

> the results support a taxonomy in which alignment functions as a binary linguistic distinction (p. 147).

However, this conclusion should be taken with appropriate caution. In the first instance, only 3 out of 5 subjects provided imitations which justify such conclusion. In addition, it should be observed that the bimodal distributions of the peak delay data can be an artefact of the durational characteristics of the experimental syllables, since the experimenters were not able to exert any control over this parameter. This possibility has certainly not been appropriately considered by the authors, since they define peak delay with respect to vowel onset and this does not take into account any durational differences of the experimental syllables that may occur within and across speakers. This in itself may have had important consequences, as will be shown below.

Let us assume for argument's sake that there is a preferred position for peak alignment as suggested by Gussenhoven, rather than two underlying categories. Let us furthermore assume that the F0 peak in this ideal position is located in the middle of the accented vowel. It follows from this that if the vowel has a duration of 150 msec, the peak should be located at 75 msec after vowel onset. If its duration is only 100 msec, the peak would have to be located at 50 msec. It can easily be seen from this that a definition of peak delay in terms of an absolute value after vowel onset does not reflect variability in the duration of the experimental vowels and this can potentially result in a bimodal distribution of the peak delay measurements: if there were a preferred peak position, the F0 peak would be systematically located later with respect to vowel onset in long vowel durations than in short vowel durations and this would be an artefact of duration, rather than the reflection of a linguistic distinction. Therefore, a relative measure of peak location which explicitly takes into account vowel duration, would have been a more useful measure.

Pierrehumbert & Steele (1987) have recognized this problem to a certain extent, in that they assert that:

> It is also important to rule out any possibility that peak delay data might be an artefact of durational differences (p. 147).

They were unable to measure the duration of the accented syllable due to the absence of a well-defined measurement point. This was partially remedied by measuring the duration of the first two syllables from [m] to [n], the argument being that durational variations in the stressed syllable should be reflected in the duration of the [m]-to-[n] stretch. On the basis of these measurements they conclude that this duration does not increase as a function of peak delay. Obviously, such a procedure can at best be a very indirect indication of the duration of the accented syllable, since a lengthening of this syllable can be quite easily accompanied by a shortening of the following unaccented syllable, thus cancelling out all relevant effects.

In conclusion, it can be said that this pitch contour imitation experiment provides some evidence that the rise-fall pattern in English is categorized by native speakers as an early and a late category. The evidence should however be taken with caution, since the measure of peak delay is defined in absolute terms, rather than relative to the syllable duration. As a result, variations in syllable duration may have been confounded with variations in peak location, which has serious implications for the interpretation of the results and any model built on them.

In Silverman & Pierrehumbert (in press), the alignment characteristics of prenuclear pitch accents are investigated in a variety of prosodic environments. The purpose of this investigation is rather different to that in Pierrehumbert & Steele (1987), in that its main aim is to examine the relationship between the structural characteristics of utterances and F0 peak alignment associated with H* accents.

In this experiment, two speakers were asked to produce a series of proper names which were chosen in such a way that the rhythmic configuration between the syllables carrying the word stresses was systematically varied. The first variable concerns the number of syllables separating the accents, which varied from 0 to 3 syllables (Ma Lemm vs. Mamalie Lemonick). A second variable is the location of the word boundary relative to the prenuclear stress. In some combinations, the prenuclear accent immediately preceded the word boundary (Ma Lemm), whereas in other combinations this was not the case (Mamalie Lemonick). In some combinations finally, there was an additional prosodic factor involved, more specifically that of stress clash. In these instances, the prenuclear stressed syllable immediately preceded a word boundary and simultaneously participated in a stress clash (Ma Lemm, Mom Lemm).

The subjects were asked to realize these utterances with a H*H+L* intonation pattern, i.e. the prenuclear accent is associated with a rise. It is followed by a stretch of high declination and terminates in a fall on the nuclear accent. In the terminology of GDI, this contour is analyzed as 10A. The informants produced five repetitions of each utterance in three different speaking rates (fast, normal and slow), which yielded a total of 360 observations. These deliveries were recorded and measurements were made of the segment durations and the location of the F0 peak corresponding to the prenuclear H* accent. The latter was measured in msec with respect to vowel onset.

The results obtained in this way are significantly affected by both speaking rate and right-hand prosodic context of the placement of the F0 peak. As far as speaking rate is concerned, it is observed that in the slow rate, the peak is characteristically located later in absolute terms than in the fast rate. One speaker, for instance, demonstrated peaks typically located at 113 msec in the fast delivery rate and at 207 msec in the slow rate. Although the parameter of speaking rate alone accounts for only a third of the variance in the data, the importance of a proportional measurement of peak delay is emphasized. This confirms our earlier observation that the results presented in Pierrehumbert & Steele (1987) should be interpreted cautiously.

Silverman & Pierrehumbert find that the influence of prosodic context on peak alignment is an important factor. This is not only because inclusion of these factors in the statistical model accounts for a larger proportion of the variance in the data (two thirds), but also because their effect on peak location is qualitatively different from that of speaking rate. It causes the F0 peak to occur earlier in the syllable:

> Both speech rate and right-hand prosodic context influence F0 peak placement, but they do so in qualitatively different ways. When a syllable is lengthened from being spoken more slowly, the peak will occur correspondingly later. In contrast, when the lengthening is induced by right-hand prosodic context, the later part of the syllable undergoes disproportionally more lengthening and at the same time the peak will occur earlier in the syllable rhyme (Silverman & Pierrehumbert, in press: 23).

Silverman & Pierrehumbert further observe that the data correspond well to those reported for nuclear accents in Steele (1986). They propose therefore a model of peak alignment in which the fine detail of peak placement is determined by a combination of length-inducing factors which exert different influences on the exact location of the peaks: syllable lengthening associated with speech rate delays the peak, whereas context-induced syllable lengthening shifts the peak to an earlier location.

Silverman & Pierrehumbert propose several possible explanations for the way in which peak alignment interacts with the prosodic context. They argue that their data support the gestural overlap/tonal repulsion hypothesis and they furthermore indicate that peak placement may in some way be related to the sonority profile of the syllable.

Gestural overlap refers to the fact that when the rising pitch movement is immediately followed by a nuclear fall, as in the stress clash condition, part of the rise may be cancelled out to allow for the complete realization of the falling movement. This could result in an earlier F0 peak placement. Tonal repulsion is closely related, but implies that the articulatory gesture associated with the rise is actually moved earlier in time to enable the pronunciation of the following pitch movement in the time that is still available. Silverman & Pierrehumbert argue that for these hypotheses to be confirmed, the location of the F0 peaks should be a function of the inter-accent distance. The relevance of this parameter could be shown statistically, but at the same time they note that it accounts for a smaller proportion of the variation in the data than their original model. This however is the result to be expected, since the gestural overlap hypothesis can only apply to contexts in which the accents immediately follow each other as in for instance the stress clash condition. When accented syllables are separated by one or more syllables, this effect can no longer apply. Nevertheless, it is clear that gestural overlap/tonal repulsion is likely to be involved to a certain extent.

A very attractive alternative explanation is given in the form of the sonority profile hypothesis, which is described as follows:

> The opening and closing gestures for the syllable give rise to an increase and decrease in sonority (where we define sonority loosely in terms of the overall openness of the vocal tract or the total impedance looking forward from the glottis). The sonority profile, or the time course of sonority for the syllable, differs between lengthening and non-lengthening environments because the closing gesture is more extended by prosodic lengthening than is the opening gesture. The F0 gesture for the accent is coupled to the entire sonority profile of the syllable (not just aligned with the vowel onset ...). The exact form of the coupling interacts with the different effects of rate and prosodic lengthening on the sonority profile, and thereby yields difference in alignment (Silverman & Pierrehumbert, in press: 5).

In this perspective, the prosodic lengthening triggers alter the sonority profile of the accented syllable in such a way that the proportional peak placement in the syllable is reduced.

The implications of the findings presented in this experiment should not be understated: the results clearly indicate that a reductionist approach to intonation, in which intonation analysis is reduced to the study of pitch phenomena only, is inadequate. Indeed, in this experiment the informants were asked to produce a single intonation pattern and nevertheless the detailed realization of this pattern is observed to vary in a highly systematic way depending on other prosodic characteristics. This emphasizes the importance of prosodic environment factors, such as speech rate and neighbouring accents, which have not been given a great deal of attention in reductionist approaches.

1.2. Evidence from Speech Perception.

Kohler (1987) investigates the alignment of rise-fall complexes, similar to those in Pierrehumbert & Steele (1987), by means of a categorical perception task. He takes his starting point in the hypothesis that in German, three categories of alignment may be relevant in the association of pitch complexes to their accented syllables. In the first category, the rise-fall complex is located on the syllable preceding the stress. In the second, the F0 peak is located in the middle of the stressed syllable. In the third category, the peak is located at the end of the stressed syllable. Kohler furthermore argues that this physical alignment difference correlates with a shift in meaning from 'established' to 'new' to 'emphatic' respectively.

In order to investigate the perceptual nature of these alignment differences, a standard categorical perception experiment was carried out, the stimuli of which were derived from the sentence 'Sie hat ja gelogen'. This utterance was read by a male native speaker of German, LPC analyzed and resynthesized with different locations of the F0 peak. These were shifted to the right from the syllable 'ge' to the syllable 'en' of 'gelogen' in steps of 30 msec, so that 11 stimuli were obtained.

The first part of the experiment consisted of an identification task. For this purpose, eight stimuli were combined with a context-setting question: 'Jetzt versteh ich das erst', which was read by the same speaker and which was also LPC resynthesized. This sentence sets a semantic context which suggests that new information is going to follow. As a result, context and semantic content conveyed by the pitch contour of the stimulus can be expected to match in a number of instances. In other cases, context and meaning do not match. These combinations were recorded on tape and presented to 19 listeners, who were asked to indicate whether context and stimulus were congruous or not.

For the discrimination task, the stimuli were combined into AX-discrimination items in such a way that items with different physical differences between the F0 positions were obtained. In one set of items, the locations of the F0 peaks were identical in both stimuli. In the second set, the peak locations differed with 30 msec. In the last set, peak positions differed with 60 msec. In addition, the stimuli in the items were presented in both AB and BA order. In the former, the stimulus with the earlier peak was presented first. In the second, the stimulus with the later peak came first. These items were presented to 39 subjects, who were required to indicate whether they could perceive any intonational differences between the stimuli of each item.

The results of the identification task show a very abrupt change in the identification function from non-matching to matching judgements. This indicates that a gradual change along the physical continuum of peak location is perceived in terms of two categories: the first category correlates with a meaning of 'established', while the second represents 'new' information. In the former category, the peak precedes the stressed vowel, whereas in the latter, the peak is located well into this vowel.

This distinction is also reflected in the obtained discrimination functions. The curve for the items with no physical difference between the stimuli shows a clear peak in the number of false alarms at the identification boundary. Those for the 1 and 2-step task have peak locations at this boundary as well. Taken together, it can be concluded from these results that the location of F0 peaks in German is perceived categorically, which in turn can be regarded as a reflection of a linguistic distinction between an early and a late alignment.

Hill & Reid (1977) conducted an AX discrimination experiment on rise alignment in English: the informants listened to a series of word pairs, in which the pitch rise of the second word occurred later than the rise in the first word of the pair. Subjects were asked to pay attention to these pitch cues and to evaluate these by giving 'same' or 'different' judgements.

The stimuli for this experiment were derived from the nonsense word /mamama/, which was synthesized on a parametric synthesizer. The stimuli were given an artificial pitch contour, the overall frequency of which was declining at 1 semi-tone per 100 msec. On this overall declination line, a pitch rise was superimposed, which had an excursion size of 4 semi-tones over 100 msec. The rise was located at the medial syllable. The stimuli were derived by shifting the rise to the right in steps of 10 msec. The discrimination items were obtained by combining the stimuli with different rise locations in discrimination pairs in such a way that the rise in the second word of the pair came later than the rise in the first word with 10, 30, 50 and 70 msec respectively.

The results of this experiment indicate that informants distinguish between three categories. The first contains rises which typically begin during the steady-state portion of the releasing consonant and extend across the release of the nuclear vowel into its steady-state portion. The rises in the second category start in the steady-state portion of the vowel nucleus and extend into the steady state portion of the following consonant. Finally, Hill and Reid provide some evidence as to the existence of a third category, the rises of which are entirely situated outside the vowel nucleus.

As far as the relevance of these conclusions is concerned, it should be mentioned that the utterance in this experiment consisted of three open syllables. Consequently, the obtained discrimination results have to be related to whether the movements are situated within the same syllable or not. The category of late movements consists of rises which are only partially located in the same syllable, since they extend across the syllable boundary into the next. The category of very late movements are entirely situated in the next syllable. If this is taken into account, it can be observed that informants in fact do not discriminate between pitch rises which are entirely situated within the same syllable and that all the movements which are located within the same syllable are regarded as instances of the same phonetic category.

A further remarkable aspect of the Hill & Reid data is that there are not more than three discrimination peaks. Since a fairly long continuum was used, there were quite a number of discrimination items which had identical rise location-conditions as the ones that are discriminated in the second syllable. The only difference was that they were located one syllable earlier (in the first syllable). In these instances however, informants do not discriminate between the stimuli. It is not immediately clear why this should be the case, since informants were instructed to listen for intonational differences in the stimuli generally. It seems possible that the attention of the subjects was drawn to the second syllable during the training session, the aim of which was to illustrate what kind of intonational differences they had to listen for. It may well have been the case that all the trial examples were instances of intonational differences related to the second syllable of the stimuli. As a result, their attention may have been directed to the second syllable, so that informants were induced to neglect melodical differences at the first syllable.

2. Pitch Movement Alignment in Dutch.

The perception of pitch movement alignment in Dutch was investigated in Collier (1972, 1975a). In Collier (1975a) an experiment was reported in which the position of a 10.5 semi-tone pitch rise on a sentence final syllable was varied from an early to a very late location in steps of 20 msec. The rise was preceded in the pitch contour by an invariant rise-fall pattern. The continuum that was obtained in this way is illustrated in figure 3.2:

Figure 3.2: Illustration of the F0 rise alignment continuum for the perception experiment in Collier (1975a).

These stimuli were recorded on language master cards and subjects were asked to sort the utterances into categories with subjectively the same pitch contour.

The most important finding of this experiment is that informants sorted the stimuli into three categories: one category consisted of early rises, one of late rises and one of very late rises. The precise location of the rises is not specified in Collier's report, but in other accounts it is defined in terms of the endpoint of the rise with respect to vowel onset. For the early category, this endpoint is situated typically at 30 msec after vowel onset, whereas the late category is located at 90 msec into the vowel. A very late rise should come as late as physically possible ('t Hart & Collier, 1975). Collier (1983) argues that these results are entirely corroborated by those obtained by Hill & Reid (1977). Although both experiments used a very similar alignment continuum, it should be pointed out that the basis for discrimination between pitch movements seems to be essentially different. While Collier's results clearly relate to the discrimination of pitch movements within the same syllable, the discrimination criteria in Hill & Reid relate to the association of movements to different syllables.

The distinction in rise alignment between early and late pitch movements in Dutch is also subject to investigation in Boves et al. (1984). Their general research aim is the development of an automatic algorithm for the transcription of intonation, but some of their results have a direct bearing on the categorization of pitch movements into different alignment categories. Boves et al. (1984) carried out an experiment in which subjects were asked to imitate sentences containing clear examples of early and late pitch movements. For this purpose, a highly trained specialist in GDI read 4 sentences which were realized with 5 different intonation contours. Three of these utterances and 3 pitch contours were used as distractors in the test. The two intonation contours on the fourth utterance served as the test items. This sentence was read with clear instances of both the hat pattern and the cap pattern, as shown in figure 3.3:

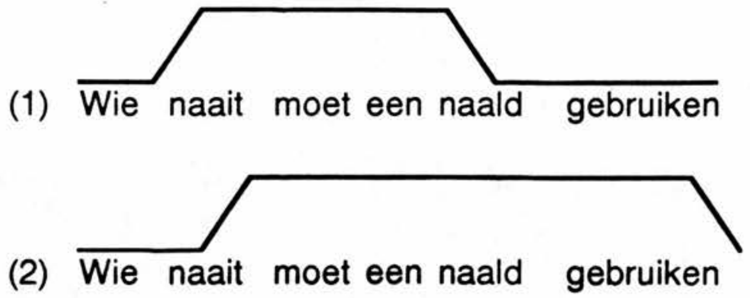


Figure 3.3: Illustration of the hat (1) and cap pattern (2) in the test items.

In (1), there is a prominence-lending rise of type 1 occurring early in the second syllable of the utterance. In (2), this syllable contains a prominence-lending pitch rise of type 3, which occurs late. In the test items, the rise onsets differ by approximately 60 msec. The second obvious difference between both intonation patterns relates to the location and identity of the terminating fall: in (1), there is a prominence-lending fall 'A' on the fifth syllable, whereas in (2), the declination line stays high until the last syllable is reached. This syllable contains a fall of type 'C'.

For the purposes of the experiment, 10 stimulus tapes were recorded, containing 5 repetitions of the utterances in randomized order. Two native speakers of Dutch imitated all 10 tapes in one session. In addition, the same speakers imitated nine of the tapes on nine consecutive working days at a rate of one tape per day. The aim of this was to obtain information about the long-term fluctuations in the production of these intonation patterns. In all cases, the imitations of the subjects were recorded on tape and F0 was measured. Finally, the alignment of the rises was established as the location of the rise-onset with respect to the start of the vowel.

From the results, it appears that the difference in alignment of the rise onset cannot be substantiated, although this distinction was clearly present in the test items:

> It is perfectly clear (...) that the obvious difference in the onset of the rises in the example utterances is absent from the imitations. The 'early-late' contrast in the examples seems to have replaced [sic] by a 'less steep, smaller excursion (type 1) - more steep, greater excursion (type 3)' contrast in the imitations. (Boves et al., 1984: 23).

At the same time, small pitch fluctuations were observed, both in the short and long term imitations (2 - 16 Hz). The authors also indicate that the early-late contrast in the example utterances was actually replaced by a contrast in excursion size. It deserves pointing out that this has to be regarded as an overstatement of the facts, since this difference in peak height was already clearly present in the two test items. The fact that this difference is observed in the imitations is consequently not surprising.

Besides this production experiment, Boves et al. (1984) also report a perception test, in which the imitations of the two test utterances were presented to a listening panel consisting of three highly trained and three more or less naïve subjects. From the report, it is not quite clear what the members of the listening panel were instructed to do with these utterances, but it appears that they had to indicate the type of rise that occurred in the second syllable of the utterances.

The results show a dramatic difference in performance between trained and untrained subjects, in that the trained informants obtained a 100% accuracy in their transcriptions. The untrained subjects only achieved a 17% accuracy. Boves et al. (1984) conclude:

> From these results we may only conclude that all imitations constituted acceptable realizations of hat or cap patterns, as all three 'trained' judges reported that they had based their decisions entirely on the contour as a whole (p. 33).

It thus seems that the ability of the trained subjects to distinguish and accurately label the type of pitch rise involved is an artefact of their familiarity with the GDI transcriptional system, i.e. this system forces them to identify the rise as 1 or 3 on the basis of the easily noticeable differences in the final portions of the contours. The grammar of Dutch intonation indeed dictates that the rise in the hat pattern is of type 1, whereas in the cap pattern, a rise 3 has to be expected. The naïve informants, who were not familiar with the GDI, could not rely on this meta-system, and were not able to identify the rises.

In order to avoid the trained subjects relying on information in the second half of the contours in making their judgements, a set of stimulus tapes were prepared, which only contained the first two syllables of all the imitations. These tapes were presented to the same listening panel, the members of which were asked to label the type of rise involved. From the results, it emerges that the most highly trained informant and the naïve subjects are not able to label the rises accurately. The two moderately trained informants now outperform the most skilled judge. Boves et al. (1984) suggest that this is because these informants may have used a different strategy in labelling the pitch movements, by relying on the peak height rather than on alignment proper.

In conclusion, the following has to be observed. If native speakers of Dutch are asked to imitate pitch contours containing movements differing by approximately 60 msec in their alignment with an accented syllable, the productions do not reflect this physical difference. The conclusion in Boves et al. (1984) that this difference is replaced by a distinction in peak height is observed to be an overstatement, since this difference was already present in the test utterances. From a perceptual point of view, it is shown that a panel of judges cannot accurately label the rises in the imitations in terms of their alignment, if the additional cue of the general shape of the intonation pattern is removed. This has no direct relevance to the distinction between early and late movements, since informants were asked to judge the movements in the imitations rather than the test items proper. In the former, any alignment distinctions were absent. The evidence is however very interesting, since it highlights that familiarity with the GDI transcription system can strongly influence judgements.

Finally, and most importantly, it has to be concluded that the results presented in Boves et al. (1984) in fact contradict a taxonomy of pitch movements into an early and a late category. If this difference is to be an essential aspect of the distinction between the hat and cap pattern in Dutch, it has to be assumed that alignment differences will be reflected in informants' imitations of these patterns. It is however clear from the experimental results that any such alignment distinctions do not materialize.

3. Pitch Movement Alignment: A Criticism.

From a general methodological point of view, the analysis of Dutch intonation has consisted of three stages. In the first instance, the stylization method has been applied to a large number of utterances. As pointed out in the previous chapter, stylization involves the construction of perceptual analogs of F0 curves in utterances by means of a minimal set of straight line segments, the requirement being that the artificial contour is perceptually indistinguishable from the original contour. The underlying aim of the method is to eliminate variability at the micro-prosodic level, which is associated with segmental effects. At the same time, it aims to preserve macro-intonational information, which can be regarded as relevant to the perception of intonation contours. The stylization method is thus specifically designed to reveal pitch information which cannot be omitted without serious perceptual consequences and as such it can be assumed to be of potential importance to the phonetic and linguistic characterization of intonation patterns.

It was also indicated in chapter 2 that the first step in close-copy stylization consists of making an exact copy of the original contour. Subsequently, the experimenter removes as much detail as possible, until a minimal specification of the artificial contour is obtained, which yields the same perceptual impression as the full specification. It is clear that in this process, the experimenter relies on two sources of information. In the construction of the close-copy, he heavily relies on visual information, i.e. the displayed original F0 curve. In the stylization process on the other hand, he uses his own perceptual judgement. As a result, a certain amount of perceptual normalization occurs, which is assumed to eliminate any segmental effects on intonation. The macro-intonation features, such as the general shape of the pitch contour, excursion sizes and slopes of pitch movements as well as the precise location of pitch movements in syllables remain present in the stylizations, since they are close-copies of original F0 curves. Thus, any acoustic variability regarding these macro-intonational features that is present in the original F0 curves will be reflected in the stylizations.

It has to be pointed out that acoustic variability concerning all these dimensions may result from three possible sources. First, there is variability of a low-level, non-systematic nature. Just like the concrete acoustic realization of -say vowels- at the segmental level may vary a great deal both across and within speakers, it can be assumed that the realization of the same underlying intonation pattern is intrinsically variant with respect to all its acoustic dimensions.

Secondly, phonetic variability of a more systematic nature is to be recognized, which may for instance originate from the larger prosodic context in which the contour is realized. The evidence for this kind of variability has been provided by Silverman & Pierrehumbert (in press), who have shown that speech rate and right hand phonetic context exert a systematic influence on the location of pitch movements in syllables.

Thirdly, there are the systematic differences between pitch contours which are indicative of underlying linguistic differences between intonation patterns.

It is clear that all these sources of variability interact with each other in a highly complex manner in the realization of pitch contours and this complex interaction is necessarily manifested in any corpus of stylized pitch contours. Thus, given a set of stylized pitch contours, it cannot be evident to which of these sources any observed pitch differences have to be attributed to. To put this more concretely: given a set of stylizations which are very similar in their general form, but differ from each other in say the precise location of the pitch rise in the accented syllable, it is impossible to decide whether these alignment differences result from low-level variability, systematic differences in phonetic context, linguistic differences between contours or a combination of these factors.

The second step in the analysis process has consisted of standardizing the pitch movements with respect to their acoustic dimensions. For this purpose, a series of experiments into the tolerance levels of native speakers was carried out in which all the acoustic dimensions of pitch movements were investigated. Concerning pitch movement alignment, Collier (1975a) found that native speakers of Dutch distinguish between three typical locations: early, late and very late.

The third step finally, has consisted of categorizing the pitch movements in the stylizations on the basis of this knowledge about the tolerance levels. For pitch movement alignment, it is concluded that differences observed in the stylizations have to be modelled in terms of three categories in order to do justice to perceptual reality. Only the intonational differences that are known not to be perceivable can be left out since these are unlikely to fulfil a communicative function.

The general method for analyzing pitch contours that was described above may seem relatively uncontroversial. Nevertheless, a number of fundamental problems are to be recognized which have potentially important consequences for the GDI. First, it has to be pointed out that the method has limited aspirations, in that it attempts to reveal pitch information which is of perceptual importance to the characterization of pitch contours. The underlying assumption was that only the perceptually relevant pitch features can be essential to the phonetic specification of pitch contours and are potentially relevant to identify the linguistic structure of intonation. Perceptual relevance was mainly determined by means of experiments into the tolerance levels of informants regarding a variety of pitch features and the knowledge obtained in these experiments was used to standardize the pitch movements in the stylizations. In order for this procedure to be valid, it has to be assumed that the perceptual tolerances established by psycho-acoustic experiments operate uniformly in all prosodic contexts and this has to be regarded as a highly questionable point of departure. The fact that native speakers of Dutch distinguish between three different positions of pitch movements in an utterance-final syllable does not automatically justify the conclusion that they are also able to perceive this difference in other prosodic environments. If it can be shown that the tolerance levels do not operate uniformly across all prosodic environments, the three-way distinction in pitch movement alignment potentially loses much of its perceptual basis, since there may be numerous contexts in which alignment distinctions may be present in the stylizations but cannot be perceived. As a result, it cannot be relevant to the perceptual and linguistic characterization in these contexts. This problem is briefly stated in 't Hart (1979a):

> (...) the perceptual tolerances which make stylisation possible at all, do not seem to be uniform, that is, to our experience, one has to be very precise on some places, whereas on other places it does not matter very much (p. 377).

From his own experience, the perceptual tolerances appear to be non-uniform. Nevertheless, the IPO methodology has always implicitly assumed uniformity in the perceptual tolerances by extending general psycho-acoustic knowledge to all contexts in which variability in the acoustic dimension are observed and categorization problems arise.

Although this aspect has never been systematically investigated, there is indirect evidence that it is unjustified to extend the perceptual tolerance levels uniformly to all contexts. In 't Hart (1981) the JND in excursion size between pitch movements was investigated. It was found that informants are more sensitive to rising excursions than to falling movements. This clearly violates the uniformity assumption. Nevertheless, the pitch movement alignment distinctions that were found in rises are applied to the classification of falls as well.

This has potentially important consequences, both for the linguistic characterization and phonetic specification of intonation patterns. From a linguistic point of view, the GDI is possibly too detailed, in that it is likely to postulate perceptual distinctions which may not be relevant to those contours. In this context, Ladd's observation makes sense, when he writes:

> By concentrating on phonetic detail in designing their representation systems, Pierrehumbert and 't Hart express differences, but they obscure similarities (Ladd, 1983b: 725).

This overconcern with detail makes it essentially impossible to arrive at insightful generalizations concerning the linguistic structure of intonation.

From a phonetic point of view, the GDI is arguably not detailed enough: certain pitch variations may not be perceivable, but can nevertheless be highly systematic. As far as pitch movement alignment is concerned, factors of speech rate and right hand prosodic context are important. Any such variations are left unaccounted for and this yields an incomplete phonetic model of intonation, in which these systematicities are left unaccounted.

The second controversial aspect is more explicitly related to pitch movement alignment and results from the IPO reductionist approach to intonation: the relationship between the location of pitch movements and the duration of the associated syllables. In the GDI, the alignment of movements is defined in terms of an absolute, standardized value in msec with respect to vowel onset:

> The position of a pitch movement in the syllable is defined with respect to the vowel onset moment ('t Hart & Collier, 1975: 241).

This procedure of measuring alignment does not address the fundamental insight that the location of a pitch movement in the syllable is an essentially relative property, i.e. relative with respect to the duration of the associated vowel. This notion of relativity appears to be of utmost importance. If a vowel, for instance, has a duration of 100 msec and the rise endpoint is located at 80 msec into this vowel, the rise is located proportionally late, that is at 80% of the total vowel duration. If the same rise is located at 80 msec in a vowel of 250 msec, it is aligned proportionally early at 32 %. In absolute terms both rises have an identical position, whereas proportionally, they are located quite differently. The IPO treatment of pitch movement alignment misses this fundamental point altogether.

It is almost self-evident that an absolute definition of pitch movement alignment can have important consequences for the validity of this aspect of the GDI, in that the postulated distinctions in pitch movement alignment can potentially just be an artefact of segmental differences in vowel duration. Two pitch movements may be aligned differently with respect to vowel onset, but due to durational differences of the associated syllables, they may be located at proportionally the same place. 't Hart & Collier (1975) for instance indicate that the endpoint of early rises typically occurs 30 msec after vowel onset. Late rises are located 90 msec after vowel onset. For both types of rises to be proportionally located at the same place, the first rise has to be located on a vowel of say 100 msec, whereas the second on a vowel of 225 msec. In both cases, the rise endpoints are at 40% of the total vowel duration. It is quite obvious that differences in vowel duration of this magnitude are commonly observed in actual speech. Therefore, it is not unreasonable to argue for the possibility that the differences in pitch movement alignment are related to variability in vowel duration instead of reflecting meaningful distinctions.

4. Conclusions.

The evidence regarding pitch movement alignment as a meaningful dimension has to be regarded as conflicting in nature. As far as pitch movement alignment in English is concerned, there is some indication that a rise-alignment is categorized in terms of an early and a late category (Hill & Reid, 1977). A more detailed investigation of the production of pitch rises in various prosodic environments (Silverman & Pierrehumbert, in press) shows that there are various factors which influence the location of a rise in the hat pattern. The most important factors are speech rate and right-hand prosodic context and their effects are qualitatively different: lengthening the syllable by slowing down the speech rate systematically delays the F0 peak, whereas right-hand prosodic context causes the peak to occur earlier.

As far as the alignment of rises in Dutch is concerned, the evidence is inconsistent. Collier (1975a) found that native speakers of Dutch are able to distinguish between three positions of pitch rise-alignment in an utterance-final syllable: early, late and very late. Boves et al. (1984) found however that any such distinction did not materialize in informants' imitations of unambiguous examples of the Dutch hat and cap patterns, the rises of which are assumed to be aligned differently.

It may be argued that the alignment distinctions postulated in the GDI are an artefact of the general methodological approach to the analysis of intonation, since it is based on three fundamental assumptions. First, there is the assumption that it is legitimate to extrapolate knowledge about the perception of pitch movements in an isolated context uniformly to other prosodic environments. Secondly, there is the expectation that variability in the stylizations is the reflection of linguistically meaningful distinctions only. Thirdly, there is the definition of pitch movement alignment with respect to vowel onset, instead of as a proportional measure with respect to vowel duration.

It has serious consequences for the Dutch model of intonation if either of these assumptions prove to be unjustified. In a linguistic perspective, it may overgeneralize the relevance of pitch features, i.e. if variability in pitch movement alignment is observed in a variety of prosodic contexts, the model is forced to recognize its relevance on the basis of a single perception experiment. It may nevertheless be not at all perceptually relevant in those contexts. In a phonetic perspective, it misses the fundamental insight that variation in pitch movement alignment may be linked to the larger prosodic context such as speech delivery rate. Taken together, the arguments presented here are strong enough to proceed to a detailed examination of the perceptual relevance of pitch rise-alignment in the next chapter.

CHAPTER 4

The Discrimination of Rise Alignment in the Dutch Hat Pattern.

0. Introduction.

It was pointed out in the previous chapters that the GDI postulates three categorically different positions of rise-alignment on the basis of discrimination data reported in Collier (1975a). In their methodology of intonation analysis, this knowledge about the perception of pitch movement alignment is used to categorize stylized pitch contours along the alignment dimension in terms of whether the movements occur early, late or very late in their associated syllables. The underlying assumption is that contours differing along the dimension of alignment potentially represent linguistic differences. Although this approach to the analysis of intonation is commendable from a phonetic perspective, the GDI is rather unambitious in the linguistic domain in that relatively few experiments have been carried out to show whether or not phonetically different contours represent different linguistic categories.

In our view, this is particularly important with respect to the dimension of pitch movement alignment, since it has been shown (Silverman & Pierrehumbert, in press) that the precise location of pitch movements in syllables is the result of a complex interaction between the abstract intonation pattern and lower level phonetic aspects such as speech rate and other length-inducing factors. It was argued in the previous chapter that this low-level variability may have persisted into the GDI, since pitch movement alignment is defined in the GDI as an absolute value with respect to vowel onset. As a result of this, any interactions between vowel duration and alignment will necessarily be reflected in the GDI. Therefore, it is necessary to evaluate the extent to which alignment distinctions can contribute to perceptual distinctions between abstract intonation patterns, in such a way that influences of alignment variation on the basis of durational differences can be ruled out. This will be done in the experiment to be reported in this chapter by means of perceptual testing of alignment differences in segmentally identical stimuli.

The specific intonation pattern of interest in this study is the hat pattern. Although the GDI assumes that the precise location of the rise is a fundamental characteristic of this pattern, it recognizes that there is a substantial amount of variability in real speech data, in that:

> (...) some pitch movements occurred earlier or later in the syllable than is specified in their definition, to the effect that blocks had to be transcribed as e.g. 1C or 3A (where the grammar invariably predicts 1A and 3C). ('t Hart & Collier, 1975: 247-248).

Although the contour 3A is observed to occur in the IPO corpus, the GDI does not recognize it as perceptually distinct from 1A, i.e. the prototypical hat pattern, but treats the contour as an exceptional variant. This is motivated by the fact that the 3A contour does not occur very often in the corpus, but in view of the general methodological approach it has to be regarded as an arbitrary decision. Since the methodology is explicitly aimed at inventorizing all the pitch contours which are of potential linguistic relevance, the occurrence of 3A contours in the corpus has to be taken into account. The methodology essentially does not permit to interpret the 3A contour as 1A, when the rise in actual fact occurs so late that it has to be standardized as a late movement of type 3, since this difference may correlate with a linguistic distinction. Any deviation from this procedure amounts to arbitraryness, unless it can be shown that informants categorize both patterns similarly. The question of how informants categorize these two contours has never been addressed and will be investigated systematically by means of a categorical perception experiment in this chapter.

The predictions related to the outcome of the experiment can be formulated as follows. If the contours 1A and 3A are regarded as perceptually distinct contours which relate to linguistically different patterns at the abstract level of the Dutch intonational system, it can be predicted that in a discrimination task, a clear discrimination peak will be found which coincides with a category boundary established in an identification experiment. In other words, the patterns can be expected to be perceived categorically. The fact that linguistic differences between contours are perceived categorically is suggested by research in Kohler (1987), which was reported in the previous chapter. If the contours under consideration do not represent linguistically distinct patterns, the alignment differences are expected not to be perceived categorically.

1. Method and Procedure.

In the description of the experimental design of this experiment, all the relevant methodological aspects are discussed in a fair amount of detail. This decision was taken since all the points that are raised here are directly relevant to most of the other experiments reported in the next chapters. This approach will enable an in-depth treatment of these methodological matters, whilst at the same time avoiding later tedious repetition.

1.1. The Experimental Paradigm.

The paradigm for this experiment is based on the experimental design that has been developed for categorical perception experiments. Categorical perception refers:

> (...) to a mode by which stimuli drawn from a physical continuum are responded to, and can only be responded to, in absolute terms. Successive stimuli drawn from a physical continuum are not perceived as forming a continuum, but as members of discrete categories. They are identified absolutely, that is, independently of the context in which they occur. Subjects asked to discriminate between pairs of such 'categorical' stimuli are able to discriminate between stimuli drawn from different categories, but not between stimuli drawn from the same category. In other words, discrimination is limited by identification: subjects can only discriminate between stimuli that they identify differently (Studdert-Kennedy et al., 1970: 234).

The basic assumption of the traditional categorical perception paradigm is that informants can only discriminate between two stimuli from a physical continuum when these are identified as representing different phonetic categories. In order to investigate to what extent this assumption holds for a particular physical continuum, informants are required to take part in an identification experiment and a discrimination task. In the identification task, subjects are presented with stimuli from a physical continuum for classification into a number of predetermined categories. In a discrimination experiment, physically different stimuli from the continuum are presented for discrimination. The results of the labelling experiment are subsequently used to predict a discrimination function, which is evaluated against the obtained discrimination function in order to determine to what extent discrimination is limited by identification. These procedures are briefly illustrated with respect to data obtained in a categorical perception experiment on vowel duration in German (cfr. Postscript Section).

The stimuli for this experiment represented a continuum of vowel duration from 280 msec to 20 msec, so that the stimuli with long vowel sounded like the German word 'Staat' (State), while the stimuli with short vowel sounded like 'Stadt' (Town). In the identification experiment, the stimuli were presented to native speakers of German, who were required to label each stimulus as either 'Stadt' or 'Staat'. In the discrimination experiment, the stimuli were presented to informants in a pairwise fashion. In all the items, there was a 30 msec difference in vowel duration between the vowel in the first and the second stimulus. It was the task of the informants to judge whether the two stimuli were 'same' or 'different'. The results of the identification experiment are given in figure 4.1:

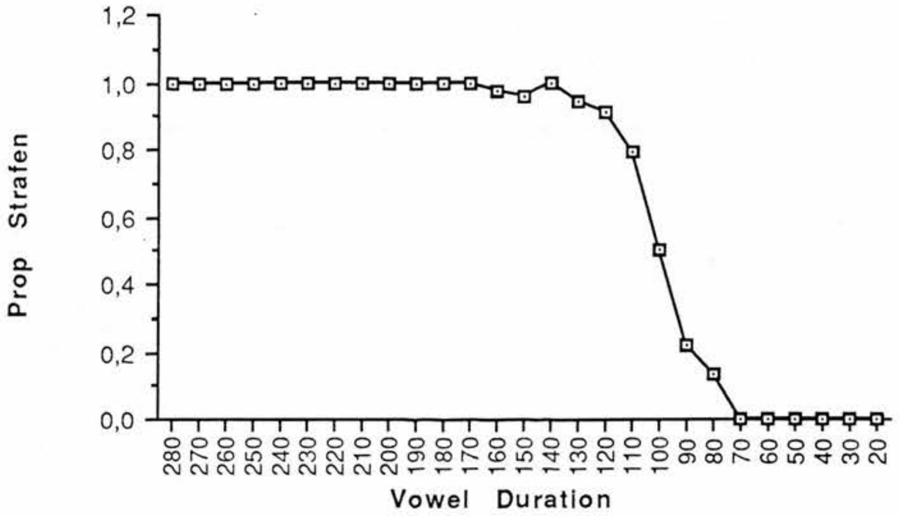


Figure 4.1: Proportion of 'Strafen' judgements as a function of vowel duration (msec).

The labelling function in figure 4.1 represents the probability of labelling the individual stimuli from the continuum as 'Staat'. It can be seen that the stimuli with long vowel duration are always labelled as 'Staat'. At some point in the continuum, this probability decreases rather suddenly to become zero for stimuli with short vowel durations: these are consistently labelled as 'Stadt'.

From this labelling function, it is possible to derive a predicted discrimination function on the assumption that informants can only discriminate between stimuli from the continuum when they are identified as representing different phonetic categories. The predicted discrimination function represents the probability of discriminating the stimuli in the discrimination items. If the two stimuli are labelled similarly in the identification task, the likelyhood that they are discriminated is very low. If two stimuli are labelled differently, the probability of discrimination is high. In this example, the probability of labelling the first stimulus and the second stimulus differently is the probability of labelling stimulus 1 as 'Staat', multiplied by the probability of labelling stimulus 2 as 'Stadt'. The likelyhood of labelling the second stimulus and the first stimulus differently is the probability of labelling stimulus 2 as 'Staat' multiplied by the probability of labelling stimulus 1 as 'Stadt'. The total probability of labelling the stimuli differently (i.e. predicted discrimination) is the sum of the two probabilities obtained above. The obtained and predicted discrimination functions are illustrated in figure 4.2:

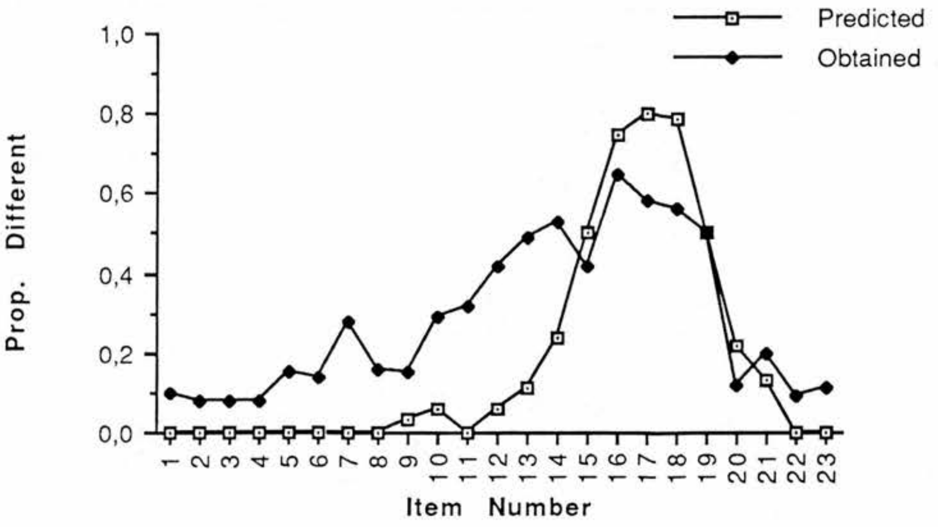


Figure 4.2 : Obtained and predicted discrimination functions for the vowel duration continuum.

In figure 4.2, it can be seen that the obtained discrimination function is better than expected discrimination over a large area of the continuum. This finding is typical if the results are analyzed in terms of the traditional Haskins model of categorical perception.

The phenomenon that obtained discrimination is better than predicted discrimination is accounted for by the so-called dual process model. This model has two parameters, i.e. phonemic sensitivity and auditory sensitivity. The first type of sensitivity is expressed as the probability of correct discrimination via the labelling process, while auditory sensitivity is the probability of being correct via the auditory process. The fact that discrimination is better than identification is accounted for by the fact that informants use two processes in discrimination, whereas they use only one in identification. MacMillan (1987) argues that extensive tests of this model have not been very successful. In one of his own applications of the dual process model, it turns out that auditory sensitivity is not constant, but varies with phonemic sensitivity. Also, auditory sensitivity is often higher than phonemic sensitivity, although the model predicts that the auditory process is only consulted when the phonemic process fails. This would suggest that informants use a nonoptimal strategy. Finally, it is indicated that the model cannot be applied to same/different tasks. In these tasks, the auditory process is never consulted since the phonetic process cannot lead to an ambiguous decision. Two stimuli are classified as either same or different and either leads to a direct response.

As an alternative to these low-threshold models, Macmillan proposes a model, which is rooted in signal detection theory. It is based on Durlach and Braida's theory of intensity perception. It is an essentially psychophysical theory in that the theoretical constructs are related to aspects of the experimental design. The model postulates that there are three sources of variance which limit performance. First, there is *sensory* variance, which originates from irreducible neural fluctuations. This source of variance is assumed to be equal for all the stimuli in an experiment.

The informants taking part in an experiment operate in trace and/or context mode. In *context* mode, each signal is compared to the overall context in which the stimulus occurs. The context variance increases as a function of stimulus range. In trace mode, informants compare stimuli with the memory trace of other stimuli and *trace* variance increases as a function of the interstimulus interval. The theory relates d' and these three sources of variance to subjects' performance in three tasks. In fixed discrimination, in which only two different stimuli occur in a block of trials, there is only sensory variance. In roving discrimination, the three sources of variance affect discrimination performance, while in identification it is only context variance that limits performance.

MacMillan (1987) observes that this theory fits the experimental data quite well. One of the observed discrepancies however, is that informants' sensitivity in identification is not strictly proportional to their sensitivity in fixed discrimination, i.e. identification seems to be good at the edges of stimulus ranges.

This is accounted for by postulating the existance of perceptual anchors: informants make judgements in context mode by comparing stimuli with well-remembered stimuli which are located at the edges of the range. Stimuli close to these anchors are identified accurately, while informants perform less well on those far away from the achors. MacMillan uses the concept of perceptual anchors to determine whether the category boundary effect in categorical perception may result from the existance of perceptual anchors inside the stimulus range. It is argued that regions of natural sensitivity are revealed by high discrimination performance in fixed discrimination tasks, since fixed performance is only limited by sensory variance. Regions with high identification or roving discrimination with respect to fixed discrimination contain perceptual anchors. This indicates that investigations of the category boundary effect should use fixed discrimination because this task minimizes the contribution of memory. In roving discrimination, the category boundary effect may arise from perceptual anchors.

1.1.1. The Experimental Task.

In the experiments to be reported, informants discrimination behaviour is systematically investigated with respect to specific pitch features in order to examine whether it is possible to establish different categories of tonal alignment. The underlying assumption to this approach is that stimuli from different categories can be expected to be markedly more discriminable than stimuli within a single category (Repp, 1984).

As has been pointed out earlier, several paradigms are in widespread use for the purpose of investigating discrimination behaviour in speech. The most common experimental design is undoubtedly the ABX(1) paradigm. In this format:

> the two stimuli being discriminated are presented, in either order, followed by a second presentation of one of them. The subject's task is to say whether the third stimulus was the same as the first or the same as the second (MacMillan & Kaplan, 1977: 453).

In this design, the possible discrimination items are <ABA>, <ABB>, <BAA> and <BAB>. This is known as the variable standard design. In the fixed standard ABX task, the potential items are <ABA> and <ABB> only.

---

(1) In a trial of a discrimination experiment, an item is presented which consists of a number of stimuli. The stimuli to be discriminated are denoted A and B. A stimulus which can be either A or B is denoted X.

Besides the ABX design, there is the 'two interval same-different' task (2IAX), in which there are 4 possible discrimination items: <AA>, <AB>, <BA> and <BB>. In this case, the informant has to indicate whether the stimuli in each item are the 'same' or 'different'.

In the 'four interval same-different' task (4IAX), a sequence of four stimuli is presented. There are eight potential items and two general item types are to be distinguished: in four items, the first two stimuli are identical, whereas the two last ones are different (eg. <AABA>). In the remaining four items, the reverse applies (eg. <BAAA>). The task is such that the informants are required to indicate to which type the items belong.

There seem to be no general rules regarding the appropriateness of any of these paradigms to the purposes of this experiment. All the paradigms share the advantage that the exact basis for discrimination need not be specified by the experimenter. On the other hand, extensive research on the task factors that are involved in categorical perception (Repp, 1984) suggests that the 2IAX paradigm is the more sensitive paradigm, which yields the most stable results when used with naïve subjects:

> (...) the results available suggest that the ABX paradigm is inferior to the AX paradigm with naïve subjects but not with experienced subjects (Repp 1984: 266).

Since naïve subjects with no formal training or experience will be used in this experiment, the 2IAX paradigm was chosen for the discrimination experiments. The factor of memory load is a further argument in favour of this paradigm. In this experiment, informants have to compare intonation contours in short sentences of up to 1.5 sec. Consequently, the 2IAX imposes the smallest demands on memory, since it involves the comparison of only two stimuli, rather than three in the ABX task or four in the 4IAX. A potential disadvantage of this roving discrimination task, in which the two stimuli to be discriminated vary from trial to trial, is the risk that discrimination peaks may result from the presence of a range internal perceptual anchor (MacMillan, 1987).

Besides the discrimination test, informants were also required to participate in an identification task. In the labelling experiment, the stimuli from the physical continuum under investigation are presented in a random order to the subjects, who are required to classify them in a number of predetermined categories. Here, the central problem is the choice of the category labels by the experimenter. This is especially problematic in intonation studies, since it is not always evident whether particular pitch distinctions correspond in a straightforward manner with a difference in meaning. This is certainly true for the pitch movement alignment continuum to be investigated in this experiment. Therefore, it was decided to use psycho-acoustic labels rather than linguistic ones. One option would have been to use the labels 'early' vs. 'late', since these are directly related to the physical dimension of alignment. This idea was however rejected, since informal experimentation has shown that naïve subjects have great difficulty in handling these concepts. As an alternative, it was decided to use the labels 'high' vs. 'low', since variation in pitch movement alignment corresponds perceptually to variation in perceived pitch height of the syllable in which the movements are located (Nabelek et al., 1970).

1.2. Stimulus Preparation Method.

In perception experiments, the quality of the acoustic stimuli used for the informants to evaluate is of fundamental importance. The main problem is the execution of perfect control over the independent variables, while at the same time obtaining natural and high quality speech output.

In this experiment, the stimuli were prepared by modified resynthesis of natural speech. In order to approximate the natural speech situation, an attempt was made to obtain as good a speech output as possible. The role of the naturalness of the experimental stimuli in discrimination performance has never been systematically investigated, but is generally assumed to play a significant role:

> Poorly synthesized stimuli may be expected to be less categorically perceived (given that they are sufficiently distinct acoustically) than good synthetic stimuli or natural speech. The reason for this is that poor stimuli make it easier for listeners to adopt auditory strategies in discrimination, whereas highly realistic stimuli may elicit a phonetic strategy (Repp, 1984: 280).

The use of the resynthesis technique that will be described here, has the additional advantage over for instance synthesis-by-rule that potential artefacts of pseudo-acoustic boundaries on pitch discrimination are avoided. The possible influence of conspicuous acoustic boundaries on pitch discrimination is indicated in Hill & Reid (1977):

> One noticeable feature of the data (...) relates to the difficulties associated with the use of synthetic speech. The discriminability peaks have a no-

> ticeable tendency to centre on what may be termed 'quasi-acoustic-boundaries' in the utterance structure (p. 343).

Speech output quality in resynthesis primarily depends on the accuracy of the acoustic analysis of the utterance that constitutes the basis of the stimuli.

1.2.1. Recording and Acoustic Analysis.

The sentence that was used in this experiment was the Dutch utterance 'Renaat is ziek' (Renaat is ill), which was read 10 times by a male native speaker of Dutch. The utterances were realized prosodically with the flat hat pattern, i.e. a prominence-lending rise 1 on 'ReNAAT', followed by a stretch of high declination, terminating in a prominence-lending fall A on 'ZIEK'. These utterances were recorded with digital recording facilities in a sound-proofed room. From these recordings, one delivery was chosen which sounded natural and fluent in all respects.
This utterance was subsequently digitized on a Masscomp MC 5500 system at F_S = 20.000 Hz. Quantization was carried out by means of a 4-channel 12 bit A/D convertor. The high sampling frequency ensured the preservation of the higher frequency components important to the perception of fricatives and stops. At the same time, aliasing was avoided by lowpass filtering the signal at F_C = 8000 Hz.
First, the durational characteristics of the experimental syllables were measured. These are summarized in table 4.1:

Segment	Frame Numbers	Duration
[Rə]	533 - 637	210 msec
[n]	637 - 663	068 msec
[ɑ]	663 - 737	185 msec
[t]	737 - 786	123 msec

Table 4.1: Durational characteristics of the experimental syllables in terms of frame numbers and segment durations.

The utterance was then analyzed by means of the API subroutine of the ILS signal processing package (ILS, 1986). This LPC-based algorithm was preferred over the

SIFT-algorithm, because the cepstrally based pitch extractor of the former is generally regarded to provide more accurate pitch estimates.

The analysis conditions for the LPC analysis were chosen with great care, since they are essential to the resynthesis of the segmental string. In this, we heavily relied on Wakita (1980). The number of filter coefficients depends largely on the frequency range of interest: an appropriate number is between Fs + 2 and Fs + 4, in which Fs is given in units of 1 Khz. Furthermore, 2 to 4 additional coefficients are required to approximate the spectral slope of the excitation source. Hence, for Fs=20,000 Hz, 28 coefficients were used in the LPC model.

Window length was set at 400 points per analysis frame, which is a suitable value for the auto-correlation method that was used, since it requires a length of at least 1.5 pitch periods. Furthermore, a hamming window was specified to suppress spectral disturbances in the high frequency region due to the edge effect of, for instance, a rectangular window.

Since the formants may change rapidly in the experimental utterance, a small window shift (ctx = 50) was chosen to obtain a smooth contour of formant frequencies. Pre-emphasis was, after some informal experimentation, set at 95%.

The result of this spectral analysis is given in figure 4.3:

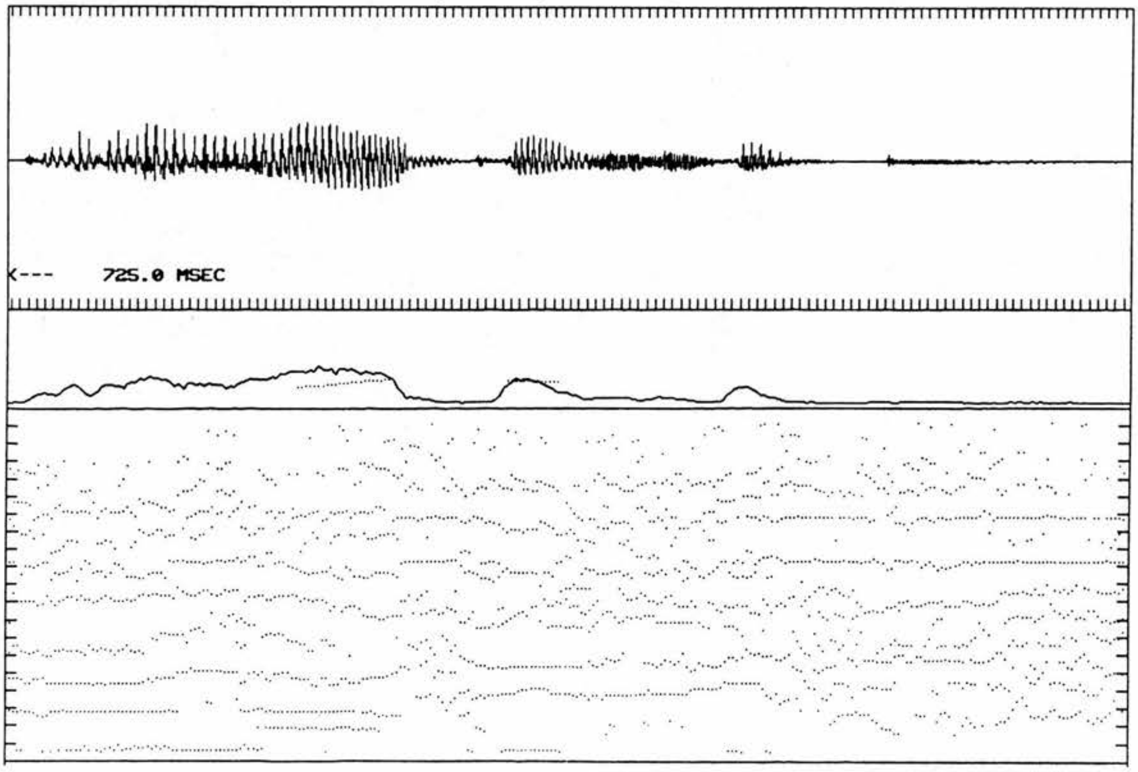


Figure 4.3: Spectral characteristics of the experimental utterance 'Renaat is ziek' after LPC analysis with conditions specified in the text. Any voicing errors were corrected manually.

After the analysis, voicing decision errors, which occasionally occurred in areas with low signal amplitude, were corrected manually. These adjustments were continuously monitored with respect to their appropriateness by resynthesizing the signal and comparing its quality to the original utterance.
The result of this procedure is an ILS analysis file, containing all the spectral information necessary for high quality resynthesis of the segmental string. All the stimuli for the experiment are derived from this analysis file by manipulating the fundamental frequency. As a result, segmentally identical stimuli can be obtained, which differ only in the parameter of F0.

1.2.2. Manipulation of F0 as the Experimental Variable.

In the next stage of stimulus preparation, the original F0 curve in the analysis file was replaced by a close-copy artificial one. For this purpose an F0 editing routine FRED (Silverman, 1985) was used. This programme enables the user to display the F0 contour of the analysis file on the screen, change F0 values of specified frames and interpolate F0 between these new values. For the interpolations, FRED provides an option between linear and spline interpolations. In all instances, spline interpolations were used, since they give smoother and more naturally sounding contours, especially in longer stretches of speech where F0 is gradually declining. The values of the turning points in the artificial contour are given in figure 4.4:

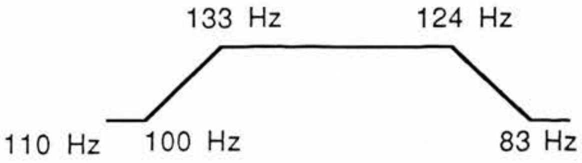


Figure 4.4: General shape of the artificial contour with frequency values of the turning points. Declination has been omitted from the schematization.

From this basic utterance, 4 stimulus types were derived by manipulating the duration of the prominence-lending rise on the experimental syllable (i.e. 50, 70, 90 and 110 msec). This yielded four intonation contours in which the slope of the rise is the only variable. These rise durations will be referred to as the 'stimulus conditions' (SC). The characteristics of the rises in each stimulus condition are summarized in table 4.2:

	Excursion Size	Duration	Frame number Start	End
SC50	5 St	50 msec	633	653
SC70	5 St	70 msec	633	661
SC90	5 St	90 msec	633	669
SC110	5 St	110 msec	633	677

Table 4.2: Summary of the rise characteristics in the different stimulus conditions.

The reasons for manipulating the slope of the rises as an independent variable will be discussed in the next section.

The actual stimuli were finally produced by shifting the rise to the right through the experimental syllable in steps of 10 msec. The remaining part of the contour was left unchanged. Consequently, there were 4 rise continua, which differed from each other only in terms of the duration of the rise. Thus 30 stimuli were obtained for SC50, 28 for SC70, 26 for SC90 and 24 for SC110. The precise location of the rises in each of the stimuli is given in table A1 of the appendix. The relative locations of the rises in the experimental syllables in each of the continua are illustrated in figure 4.5 on next page:

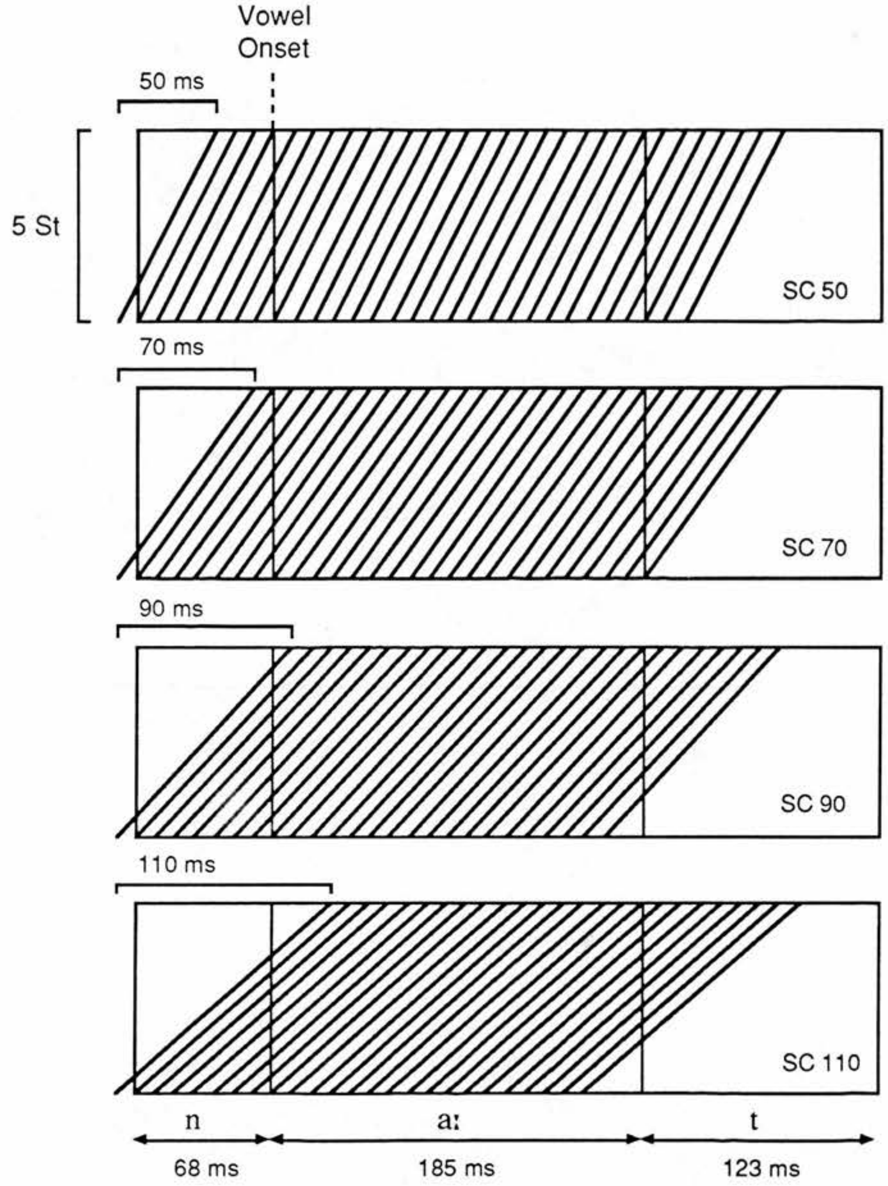


Figure 4.5: Schematic representation of the range of variation in pitch movement alignment in the four stimulus conditions. Rises have an excursion size of 5 semitones and durations of 50, 70, 90 and 110 msec respectively. The number of stimuli obtained is 30 for SC50, 28 for SC70, 26 for SC90 and 24 for SC110.

After F0 manipulation, the resulting ILS analysis files were resynthesized by means of the ILS SNS-subroutine, which is a pitch asynchronous synthesis algorithm. The filter order for the unvoiced frames was set at 28. The overall amplitude of the stimuli was scaled down to 75% in order to avoid distortion of the speech signals.

1.3. Construction of the Discrimination Items.

The stimuli obtained by above-mentioned procedure were subsequently combined into 2IAX discrimination items. The variable which is most directly related to discrimination between two stimuli in an item is the size of the physical difference between them (Repp, 1984). Careful control of the spacing between stimuli is important. If the physical difference is too small, the result will be a flattening of the discrimination function. If it is too big, a general increase in the discrimination function will be obtained, i.e. the so-called 'ceiling effect' (Schiefer & Batliner, 1988).

Following Hill & Reid (1977), 5 different gap sizes between the rise onset frames in both stimuli were used, i.e. 30, 40, 50, 60 and 70 msec. By doing this, the traditionally accepted values for the just noticeable difference of pitch movement alignment were maintained (Collier, 1983). The general structure of the items is illustrated in figure 4.6:

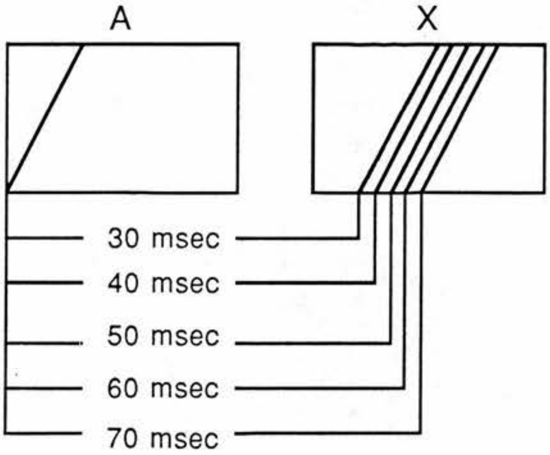


Figure 4.6: Illustration of the item structure. The gap between the rises in each stimulus is defined in terms of the onset frame of each rise.

At this point, it should be clear why the slope of the rises is manipulated as an independent variable. As a result of the combined effect of the slope of the rises in the stimuli and the gap size, the physical F0 overlap between the rises in the stimuli of the discrimination items varies. In rise duration 50 msec with a gap size of 50 msec, there is no overlap between the two pitch rises in the items. In rise duration 110 msec with the same gap size, the movements in the items overlap with 60 msec. Although there is no direct evidence, it seems reasonable to assume that the latter items might be more difficult to discriminate than the former.

In addition, two presentation orders of the stimuli in the items were used. This counterbalancing of the order of presentation is required (Pollack & Pisoni, 1971), since it has been shown in a variety of experiments that items in one presentation order are generally better discriminated than those in the other order of presentation (Batliner & Schiefer, 1988). These orders will be referred to as the AB-order and the BA-order. In the former, the stimulus with the earlier rise was presented first, while in the latter, the stimulus with the later rise came first. Henceforth, gap size and presentation order will be referred to as the 'item conditions' (IC).

This procedure of item construction yielded a total of 880 discrimination items (i.e. 440 for each presentation order). Their characteristics are summarized in table 4.3:

	Slope				
	50	70	90	110	msec
Gap Size		No. of items			
30 msec	27	25	23	21	
40 msec	26	24	22	20	
50 msec	25	23	21	19	
60 msec	24	22	20	18	
70 msec	23	21	19	17	

N= 440 items x 2 presentation orders = 880 items

Table 4.3: Summary of the item characteristics, with a specification of the number of items in each condition.

As pointed out in section 1.1.1, a 2IAX discrimination experiment traditionally includes <AA> or <BB> items, in which the physical characteristics of the stimuli are identical. The presence of these items is required in order to be able to calculate informants' response bias (Wood, 1976). Such items are not included in this experiment, because this omission reduces the total number of items in the test by a factor of two. This should cause no serious problems, since Ainsworth & Lindsay

(1986) have shown that it is possible to obtain good discrimination functions, even in the absence of these reference items.

1.4. The Test Tapes.

1.4.1. The Discrimination Task.

One of the problems in compiling the test tapes is the possible occurrence of sequence effects (Woodworth & Schlosberg, 1954), i.e. the fact that discrimination items are not discriminated in isolation, but relative to the preceding and following items. From this point of view, it is often felt unsatifactory to randomly intermix the discrimination items in the same experimental run with regard to all variables under investigation. The reason for this is that mixing items with unequal physical differences in for example gap size may exert a negative influence on the discrimination of the items with the smallest differences in gap size. If informants are first confronted with items with large gap sizes which are easily discriminated, their discrimination criteria are likely to be based on those items, with the result that the differences in the items with smaller gap sizes are obscured and are less well discriminated. In other words, randomization across experimental variables may lead to an underestimation of the informant's sensitivity:

> The effect of such randomizing in psychophysical tasks is often found to be a decrement in performance relative to that which could be obtained if these variables were held fixed (Macmillan & Kaplan, 1977: 462).

Two solutions to this problem present themselves. Ideally, informants should be presented with different randomizations across the same experimental variabels. This probably gives the most accurate estimate of true discrimination performance, since each discrimination item is tested in different contexts. As a results, the effect of context on discrimination can be averaged out. Due to the limited availability of informants, it was not possible to take this approach.

An alternative procedure is to present informants with blocks of items which are ordered in terms of increasing or decreasing gap size, i.e. not to randomize in terms of gap size. As a result, the sequence of presentation is fixed. although sequence effects cannot be avoided, they are kept constant. The precise nature of this effect is not quite clear. When the items are presented in such a way that gap size progressively decreases, it is likely that items with small gap size are underdiscriminated, since informants' discrimination criteria are likely to be based on the items with large gap sizes. When the subjects are presented with the smaller gap sizes first, it is possible that they lose interest in the task and underperform on the task as a whole. Although the ideal experimental design would have been one in which informants hear different randomizations, it was decided to present all the items in blocks with increasing gap size.

Four test tapes were recorded for the discrimination task of this experiment, one of which was used in each experimental session. The first tape contained the items with SC50 and SC70 msec in AB presentation order. The second tape consisted of items with SC50 and SC70 msec in BA order. The items for SC90 and SC110 in AB order were recorded on the third tape. The fourth tape contained the items for SC90 and SC110 in BA order. On each tape, the items were sequenced in blocks according to their gap size, in the sense that the items with the smallest gap were presented first. The rationale for separating out presentation order in different experimental sessions was that other discrimination experiments (Kohler, 1987) show that there is a marked difference in the ease of discrimination in both presentation orders.

On the tapes, the inter-stimulus interval in the discrimination items was 500 msec, whereas the inter-item interval amounted to 2,500 msec. There were no identifiers between the discrimination items, but after presentation of 10 items, the informants heard a short tone in order to enable them to orientate themselves in the task.

1.4.2. The Identification Task.

For the purpose of the identification task, a fifth tape was recorded which contained all the individual stimuli in a random order. The inter-stimulus interval was 2,500 msec.

1.5. The Informants.

For the identification task, 39 undergraduates were recruited from the Department of Business Studies of the University of Antwerp (Belgium)(2). For the discrimi-

---

(2) Contrary to popular belief, Belgium is not a French-speaking country. There are three officially recognized languages: Dutch, French and German. According to recent statistics (Belgoscopie, 1989) 60 % of the population speaks Dutch, 38 % French and 2 % German.

It should also be mentioned that sometimes the term 'Flemish' is erroneously used to refer to the variety of Dutch spoken in Flanders. This has been the cause for severe nation task, 4 groups of students (total N = 96) of the 'Economische Hogeschool Limburg' (Belgium) took part on a voluntary basis. All informants were native speakers of Dutch and no one took part in more than one experimental session.

It was decided to work with linguistically naive subjects as far as intonation is concerned. The main reason for doing so was that it has been shown that familiarity with particular approaches to intonation strongly biases the informants: it is clear from Boves et al (1984) that trained informants respond in a very different fashion than naive subjects, in that trained informants were shown to hear intonational differences in terms of the transciption system with which the informants are familiar. The use of naive subjects should avoid any such effects.

1.6. Procedure.

For all tasks, the informants were seated in a quiet language laboratory, in which the volume-level of the headphones could be individually adjusted. The subjects were told that they were going to take part in a perception experiment on Dutch intonation. The meaning of the term 'intonation' was explained in terms of 'speech melody' and was illustrated by several examples.

In the *discrimination* tasks, the informants were asked to compare the speech melody of the first word in each stimulus of a discrimination item (i.e. 'Renaat'), and were asked to indicate whether they could hear any melodic differences. They were urged to give their own personal opinion and were reassured that the task was not a form of multiple choice test with correct or wrong answers. Subjects' judgements were given on a scoring sheet with each item to be rated as 'same' or 'different'. Before the start of a session, the informants listened to 10 practice items in order to get accustomed to the task and take the opportunity to adjust the volume of the headphones to a comfortable listening level. Then the tape was stopped for further questions. The total duration of each session ranged between 20 and 45 minutes. At approximately every 10 minutes, a short break was provided in order to reduce fatigue effects.

The procedure for the *identification* task was similar to the one described above, except for the fact that informants had to judge whether the pitch on the experimental syllable of each stimulus was 'high' or 'low'. Before the start of the labelling

---

terminological confusion. Suffice it to point out here that the correct view of the situation is that the standard language of Flanders is Dutch.

session, the informants were presented with 20 clear examples of each category to familiarize them with the meaning of these labels.

2. Results.

In the discrimination experiments, a total number of 20,840 observations were obtained, i.e. 4,560 for the items with rise duration 50 and 70 msec in AB-order, 5,280 for 50-70 msec BA, 5,000 for 90-110 msec AB and 6,000 for 90-110 msec BA. Observations for discrimination items in 110 msec AB are missing as a result of technical problems with the tape recorder during the experimental session. The raw discrimination judgements for the items in the different conditions are given in the appendix (tables A2-A9). In the identification task, 11,076 observations were acquired: these judgements are given in table A10 of the appendix. In the following sections, these results will be analyzed in detail.

2.1. Statistical Analysis.

The main aim of this experiment is to investigate the effect of the alignment of the rise in the experimental syllable on subjects' discrimination performance. The control variables in the experimental design are presentation ORDER of the stimuli in the discrimination items, GAP size between the pitch movements in the stimuli and SLOPE of the pitch movements. For this purpose, an analysis of variance was carried out, which enables a comparison of mean item discrimination in all the experimental conditions simultaneously and assess the significance of any differences between the mean judgements. In this analysis, there are 4 independent variables: ORDER (AB vs. BA), GAP (30, 40, 50, 60 and 70 msec), SLOPE (50, 70 and 90 msec) and ALIGNMENT. Part of the results for slope 110 msec are missing as a result of a breakdown of the language laboratory. As a result, the design of the experiment is no longer factorial. Therefore, it was decided not to include the discrimination results for slope 110 msec in the statistical analysis in order to make de design factorial again in order to be able to obtain an insight in the possible interactions between the independent variables.

A basic problem regarding the design of this experiment is that there was only one observation per subject for each combination of experimental variables. In order to increase the number of observations for each subject, it was decided to quantify the variable ALIGNMENT in 5 different levels from 'early' to 'late' location in the experimental syllable so that meaningful proportions of 'different' judgements could be obtained for all the experimental conditions. These proportions were subsequently taken as the dependent variable. The analysis of variance was carried out by means of the SPSS-X software package and the results are summarized in table 4.4:

```
                    ANALYSIS OF VARIANCE

                    DIFF    Proportion of Different Judgements.
              by    GAP     Gap Size.
                    SLOPE   Slope of Pitch Movements.
                    ORDER   Presentation order.
                    ALIGN   Location of pitch movement in syllable.

Source of               Sum        DF    Mean          F        Sig.
Variation               Squares          Square

Main Effects             15.590   11    1.417        15.997   .000
  ORDER                   8.329   1     8.329        94.006   .000
  SLOPE                   3.256   2     1.628        18.375   .000
  GAP                      .991   4     .248          2.798   .025
  ALIGN                   2.986   4     .746          8.425   .000

2-Way Interactions       13.078  42     .311          3.515   .000
  ORDER   SLOPE            .628   2     .314          3.547   .029
  ORDER   GAP             1.059   4     .265          2.989   .018
  ORDER   ALIGN           8.119   4     2.03         22.909   .000

None of the other interactions are significant.

Explained                35.889149      .241          2.719   .000
Residual                291.7453293     .089
Total                   327.6343442     .095

4826 cases were processed.
```

Table 4.4: Result of a four-way analysis of variance by means of the SPSS-X software. The independent variables are ALIGNMENT, SLOPE, ORDER and GAP. The dependent variable is the proportion of different judgements.

It appears from this analysis that all main effects are statistically significant at. As a result, it seems possible to conclude that the independent variables ALIGNMENT, SLOPE, GAP and ORDER each have a significant effect on discrimination performance. In order to provide a clear picture of the relationship between each independent variable and average discrimination, the mean proportion of different judgements in the four conditions is plotted in figures 4.7-4.10:

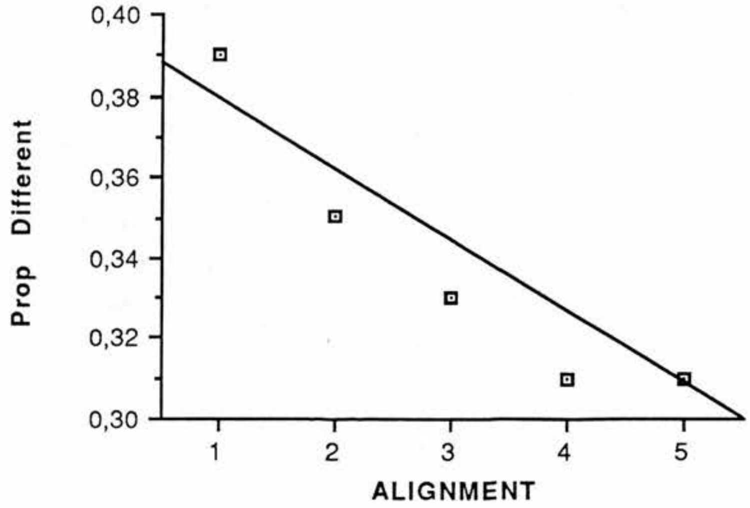


Figure 4.7: Mean proportion of different judgements as a function of pitch movement ALIGNMENT in the experimental syllable. The number of observations for each data point is 690.

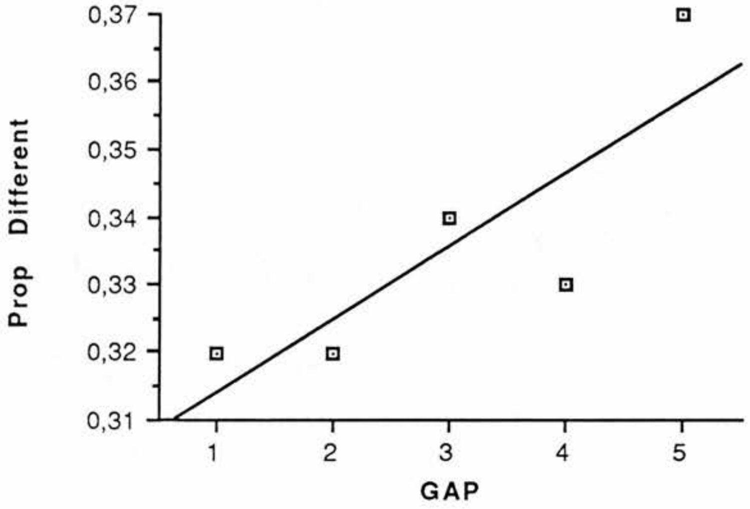


Figure 4.8: Mean proportion of different judgements as a function of GAP size, with a best fitting curve through the data points. The number of observations for each data point is 690.

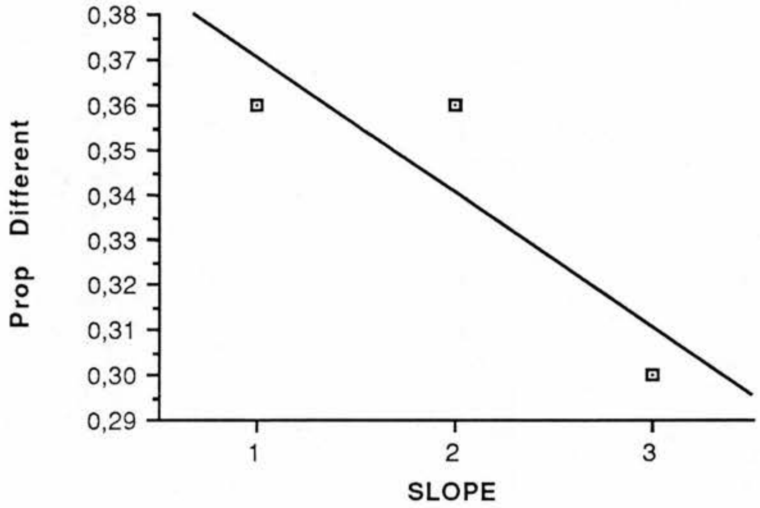


Figure 4.9: Mean proportion of different judgements as a function of SLOPE, with a best fitting curve through the data points. The number of observations for each data point is 1025, 1019 and 1399 respectively.

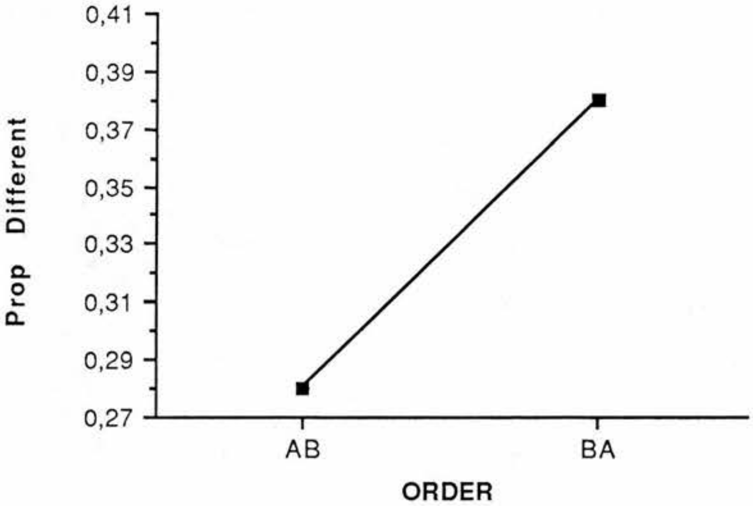


Figure 4.10: Mean proportion of different judgements as a function of presentation ORDER. The number of observations for each data point is 1593 and 1850 respectively.

A first general observation concerning these data is that the overall discrimination is very low: in neither case does average discrimination exceed 34%. This indicates that the majority of informants are not able to discriminate between the stimuli in the discrimination items. Nevertheless, some sets of items are clearly better discriminated than others.

The best-fitting curves through the data points in figures 4.7-4.10 give an indication of the general trends in the relationship between each variable and average discrimination. In figure 4.7, the relationship between ALIGNMENT and discrimination is plotted. This turns out negative, i.e. discrimination decreases as the rises are aligned later in the experimental syllable. In figure 4.8, the relationship between GAP and discrimination turns out positive: i.e. as gap size increases, the stimuli are better discriminated. This is consistent with the expectation that larger physical differences are easier to perceive. The relationship between the variable SLOPE and discrimination shows a clear negative trend (figure 4.9). This suggests that discrimination progressively deteriorates as the slope of the rises in the stimuli becomes less steep. This perceptual effect can be accounted for in terms of the physical overlap between the F0 rises in the discrimination items, in that this overlap is greatest in the items with the most gradual slopes. The trend for presentation ORDER in figure 4.10 is positive, which means that the stimuli are better discriminated in the BA (late-early) order.

It should, however, be mentioned that this interpretation of the results, as it stands, may not be entirely realistic, since all the two-way interactions involving the variable presentation ORDER in table 4.4 are significant. This requires thorough investigation, before drawing final conclusions about the relevance of the main effects themselves. Indeed, in establishing the statistical significance of for instance the main effect of ALIGNMENT, the ANOVA technique confounds the other variables ORDER, SLOPE and GAP with ALIGNMENT. As a result, the significance of ALIGNMENT may be caused by the variable with which it interacts and consequently claims about the variables that enter into the interaction may have to be modified (Hatch & Farhady, 1982). In order to make the interpretations of these interactions more insightful, they will be presented in graphical format.

The effects of GAP and ORDER on discrimination are illustrated in figure 4.11:

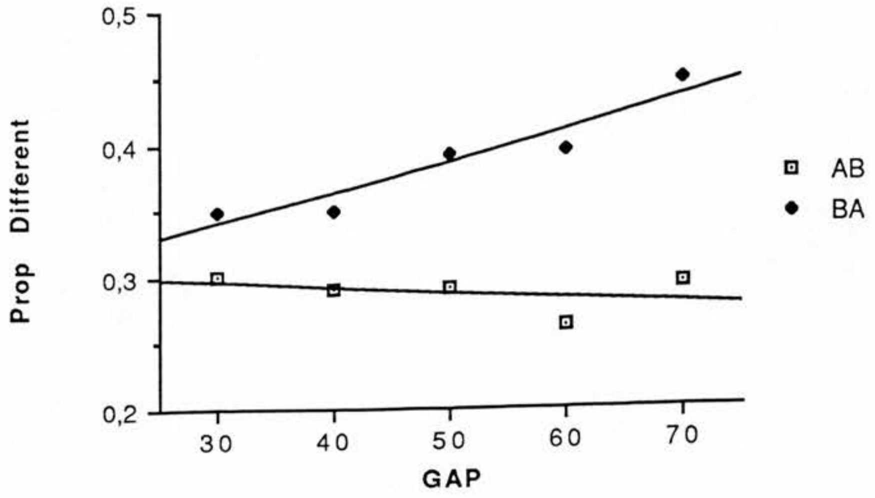


Figure 4.11: Proportion of different judgements as a function of GAP size in the AB and BA presentation order, with best-fitting lines through the data points.

Figure 4.11 provides a clear picture of the relationship between GAP size and presentation ORDER. It can be seen that overall discrimination of GAP size in the two presentation orders is essentially different, in that all the items in BA order are better discriminated than those in AB order. This is the case for all gap sizes, although this effect is more outspoken in the items with larger gap sizes. The best-fitting lines through the data points in both presentation orders also provide an indication of the relationship between GAP size and discrimination performance. It is evident from figure 4.11 that discrimination improves as a function of gap size in the BA presentation order, whereas discrimination in AB order remains at a virtually constant level irrespective of GAP size. This representation of the results requires our initial conclusions about the relationship between GAP size and discrimination to be modified. It was suggested earlier that the data show a positive correlation between GAP size and discrimination, which can now be seen to be valid for the BA presentation order only. Our earlier conclusion about better discrimination of the items in BA order is confirmed by these results.

The relationship between SLOPE and presentation ORDER is illustrated in figure 4.12:

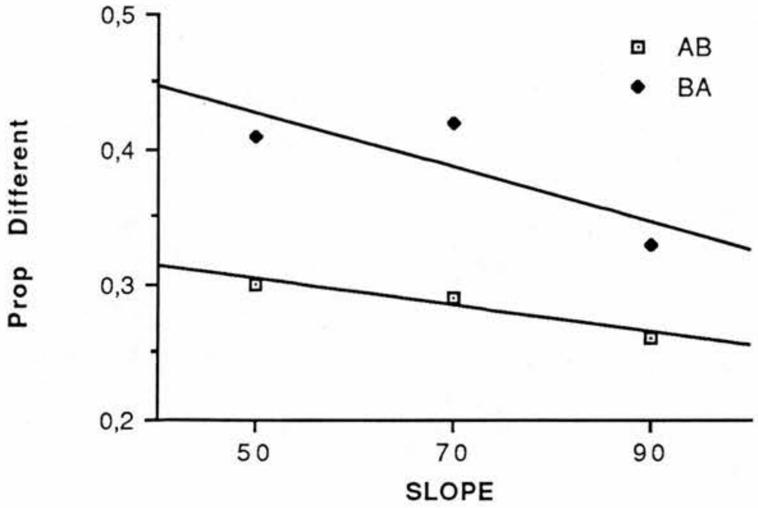


Figure 4.12: Proportion of different judgements as a function of SLOPE in the AB and BA presentation order, with best-fitting lines through the data points.

Figure 4.12 indicates that the BA order is better discriminated than the AB order. It can also be seen that discrimination performance deteriorates as the slope of the rises in the stimuli becomes shallower. This effect is most outspoken in the BA order, which suggests that it is only in the BA presentation order that there is a clear negative relationship between SLOPE and discrimination, while any such correlation cannot be maintained for the AB order.

The relationship between ALIGNMENT and ORDER is illustrated in figure 4.13:

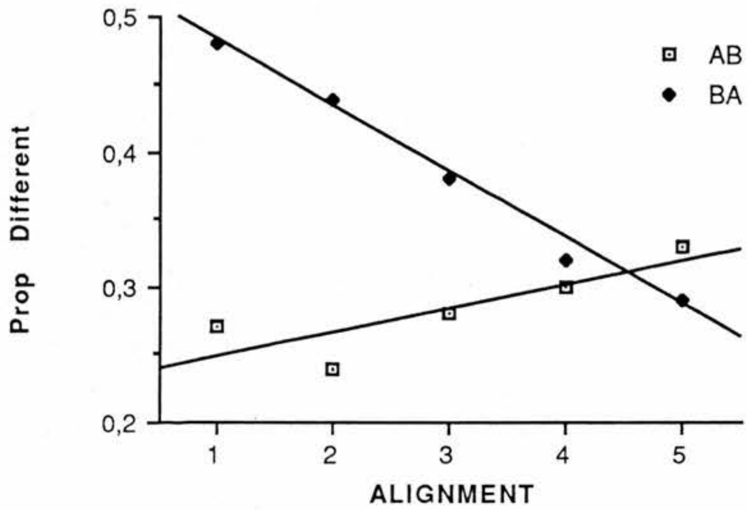


Figure 4.13: Proportion of different judgements as a function of ALIGNMENT in the AB and BA presentation order, with best-fitting lines through the data points.

In figure 4.13, it is clear that discrimination progressively decreases in BA order when the rises are located later in the experimental syllable. In AB presentation order, the effect is reversed: when the rises are aligned later, discrimination increases slightly.

From these observations, it can be concluded that informants respond differentially to all experimental variables as a function of presentation ORDER of the stimuli in the discrimination items. In early-late presentation order, there is only a very wek relationship between GAP, SLOPE, ALIGNMENT and discrimination. In the late-early order of presentation, the effect of increasing SLOPE becomes more outspoken negative, whereas the effects of GAP size and ALIGNMENT are reversed as compared to AB presentation order. This suggests that the significance of the main effects in the analysis of variance is fundamentally caused by the presentation order of the stimuli in the discrimination items.

It has to be pointed out finally, that the 2-way interactions between SLOPE-ALIGNMENT and GAP-ALIGNMENT are not significant. This indicates that the discrimination of rise alignment is essentially identical for all the gap sizes and all the slopes.

2.2. The Variable Rise Alignment.

In this section, the obtained results are further analyzed with specific reference to the variable of rise ALIGNMENT. For this purpose, the discrimination data are transformed in such a way as to eliminate the other variables from the design. This is achieved by considering each discrimination item in terms of its mean rise position, a technique that was successfully used by Hill & Reid (1977). Subsequently, the data from the identification task will be considered and their relationship with the discrimination data is discussed.

2.2.1. Discrimination Data.

Inspection of the individual discrimination curves, derived from the raw scores in all the experimental conditions, does not reveal a great deal of insightful information concerning the perception of pitch movement alignment. Let us take the raw discrimination curves for stimuli with slope 50 msec in AB order (figure 4.14) as a starting point:

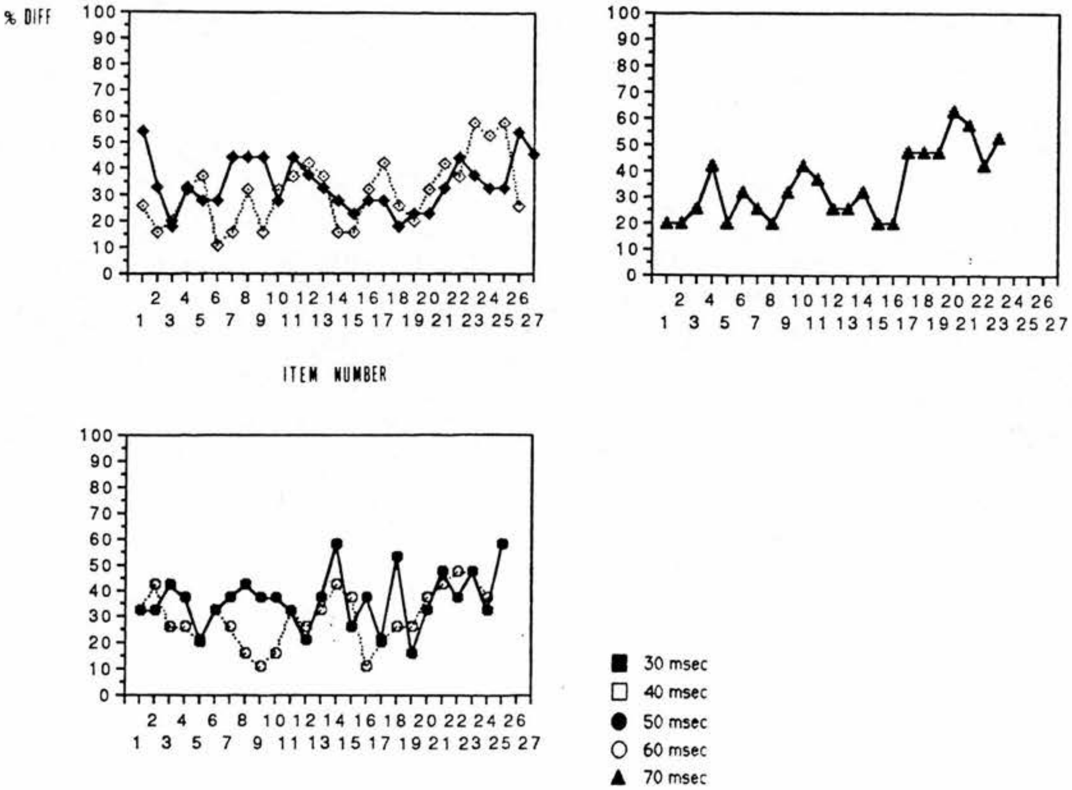


Figure 4.14: Raw discrimination curves for stimuli with slope 50 msec in AB presentation order.

If we consider the discrimination functions for the different gap sizes in figure 4.14, the results turn out to be quite inconsistent. For some gap sizes, there are clear discrimination peaks, whereas they are totally absent in the functions for other gap sizes. Moreover, if peaks do occur, they show up at entirely different places in the continuum. Thus, these data suggest that no clear and consistent category boundaries can be found, especially if it is taken into account that this inconsistency is characteristic of the discrimination data in all the other stimulus conditions. It should be noted that there are only 19 observations for each item in figure 4.14. Therefore, the data were transformed in such a way as to increase the number of observations for each item, while at the same time preserving potentially relevant effects of the different variables involved.

It was argued earlier that the variables SLOPE and GAP are only marginally relevant in accounting for discrimination differences due to their significant interactions with presentation ORDER of the stimuli. The latter was observed to be the more important factor. In addition, no significant interactions between GAP-ALIGNMENT and SLOPE-ALIGNMENT were found. This suggests that rise alignment is perceived similarly across all slopes and gaps. As a result, it is legitimate to eliminate the variables SLOPE and GAP from the experimental design. ORDER and ALIGNMENT on the other hand are of primary interest. Since it is necessary to preserve any effects of ORDER and ALIGNMENT, the data are transformed by calculating the Mean Rise Position (MRP) for each discrimination item. Mean Rise Position is defined in terms of rise endpoint(3). In terms of rise endpoint, MRP for each discrimination item can be obtained by adding up the endpoint frames of the rises in both stimuli of an item and dividing this sum by two. The essence of this technique was proposed by Hill & Reid (1977), who were confronted with a similar data reduction problem. Thus, the endpoint frame of the Mean Rise was calculated for each discrimination item and the observations for all the discrimination items with the same MRP were averaged. The data transformation suggested here enables us to eliminate the variables of SLOPE and GAP from the experimental design, while preserving the effects of the rise ALIGNMENT in the syllable and presentation ORDER. A further advantage of this technique is that the number of observations per item is increased. This data transformation procedure yields the discrimination function in figure 4.15:

---

(3) The data transformations discussed here were also carried out in terms of rise onset. The obtained results are entirely parallel to these for rise endpoints.

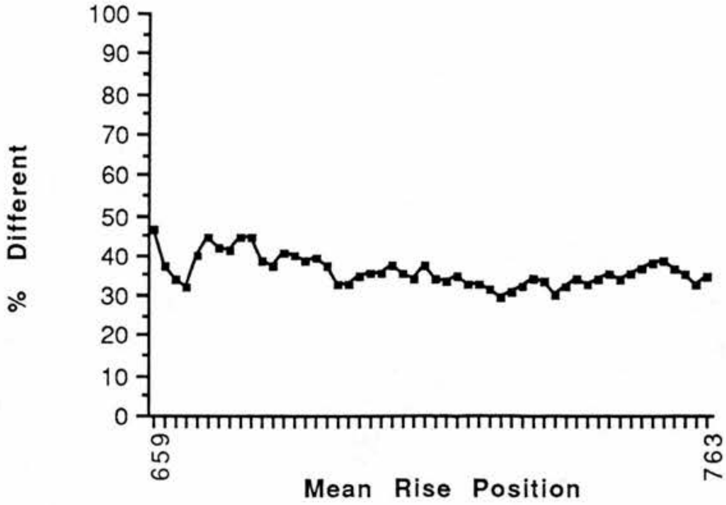


Figure 4.15: Average discrimination curves as a function of mean rise position defined in terms of rise endpoint for the total population. Total number of observations is 20,840.

Concerning the relationship between the location of pitch movements in the experimental syllable and discrimination, it is clear that discrimination remains at a fairly constant level throughout the continuum: average discrimination is 34%. It can be observed that there are no significant peaks in the discrimination curve. Therefore it has to be concluded that the existence of any categorical boundaries in the alignment of pitch rises cannot be substantiated on the basis of the discrimination data.

A possible objection to this conclusion could be that the absence of discrimination peaks in figure 4.15 has resulted from the averaging procedure that was used in the data transformation. This, however, is highly unlikely, since it has been shown convincingly by Hill & Reid (1977) that the presence of consistent discrimination peaks in the raw data remains present in the averaged curve.

2.2.2. Identification Data.

These data were obtained in a separate experimental session, in which the informants were required to identify each of the stimuli in terms of the labels 'high' and 'low'. The typical result to be expected if there is a categorical boundary is a Z-shaped identification curve, the cross-over of which coincides with the discrimination peak in the discrimination function.

The identification function is given in figure 4.16:

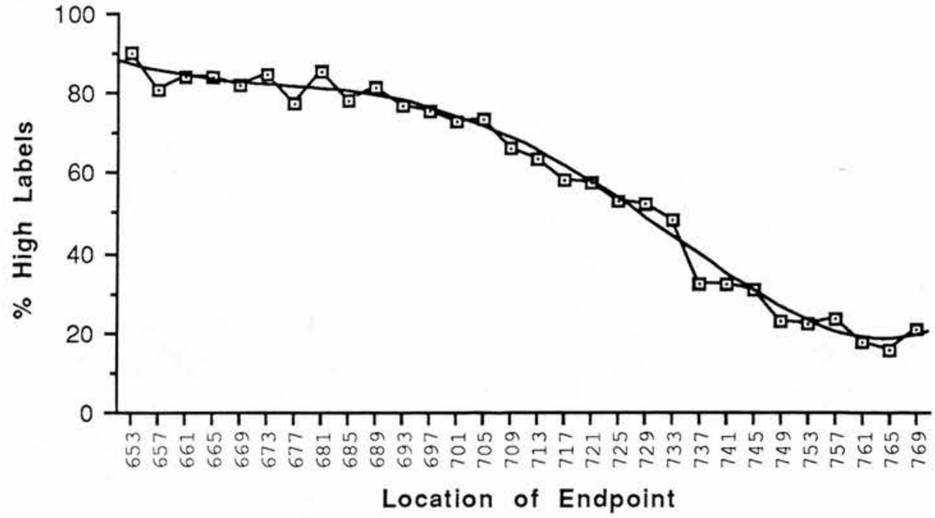


Figure 4.16: Percent High judgements as a function of the endpoint of the rise in the stimuli, which is indicated in ILS frame numbers, with best-fitting curve.

The curve in figure 4.16 does not have the Z-shape which is characteristic for categorical perception. The curve declines very smoothly as the pitch movement is shifted to the right through the experimental syllable. This labelling function thus reflects a very gradual change in subjects' identification and this is totally unlike what could have been expected if there were a categorical boundary in the pitch movement alignment continuum. Thus, this result is entirely in agreement with the absence of significant peaks in the discrimination data.

2.3. The Variable ORDER.

From the analysis so far, it has emerged that the only relevant independent variable with a significant effect on discrimination is that of presentation order of the stimuli in the items. At this stage, the effect of this variable is singled out from the overall discrimination curve that was obtained on the basis of MRP. The effect of presentation order, which will henceforth be referred to as the order effect (OE), can be quantified as the proportion of discrimination judgements in the AB order minus the proportion of judgements in the BA order. This quantification of the OE as a function of mean rise position is illustrated in figure 4.17:

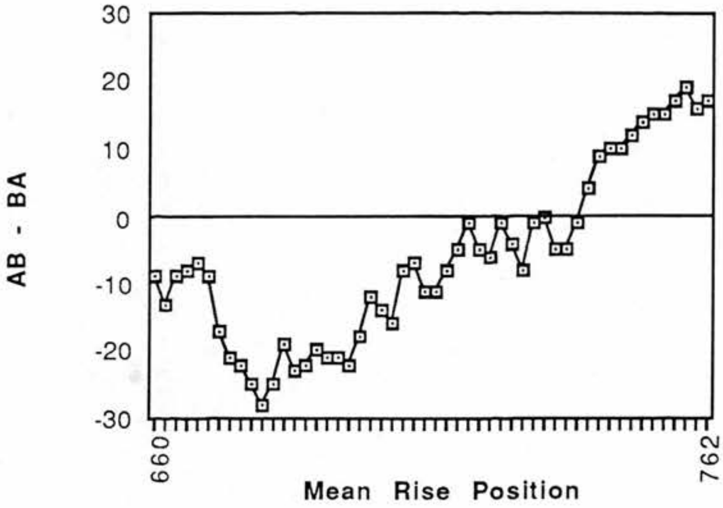


Figure 4.17: Quantification of the order effect (AB-BA) as a function of mean rise position for the total population. MRP is defined in terms of rise endpoint. Total number of observations is 20,840.

This quantification of the order effect reveals that OE is biggest at the extremes of the stimulus continuum. A further remarkable aspect of these data is that the precise nature of the effect changes throughout the continuum. In the left-hand side of the continuum, OE is negative, which means that the BA order is better discriminated than the AB order. In this area of the continuum, subjects are better at discriminating items in which the stimulus with the later rise location is presented first. Proceeding to the right in the continuum, the OE gradually disappears, in order to become positive in the right-hand side of the continuum. The latter means that the AB order is better discriminated than the BA order.

These results clearly indicate that the OE is much more complicated than suggested earlier, in that the BA order is not uniformly better discriminated than the AB order. Rather, the BA order is better discriminated in the left-hand side of the continuum. Subsequently, OE disappears and reverses in the right-hand side of the continuum, where AB-items are more often discriminated than BA-items.

2.4. Subject Groups and Discrimination Criteria.

A possible disadvantage of the experimental design is the fact that four different groups of subjects have taken part in the discrimination experiments, each of which has judged a subset of the total number of stimuli. This decision was taken on the basis of purely practical considerations, i.e. the limited availability of informants and the need to reduce listening time. So far, the 96 informants have been treated as if they constitute a single homogeneous group. This may be unjustified, since it is possible that the informants in each of the groups have consistently used different discrimination criteria. Therefore, it was decided to investigate subjects' individual response patterns in detail in order to acquire information about the response behaviour of the individual subjects. For this purpose, the technique of 'response coincidence analysis' is used.

2.4.1. The Active Role of the Subject.

In discrimination experiments of the kind detailed above, an attempt is made to establish a relationship between values in a physical continuum (viz. F0 alignment) and discrimination responses elicited from the informants. An important factor to be considered is the role of the participating subject. If it is assumed that the subject is entirely passive or is undergoing the discrimination items, variations in subject responses can be explained legitimately in terms of variations in the stimuli. It is this view of the subject as a black box, which has been adopted in much of the research into the categorical perception of speech. In such a perspective, the subject is regarded as a stimulus-driven organism. Pollack & Pisoni (1971: 300) observe that this assumption "does not admit to modification of the criterion for responding 'same' or 'different'" and thus disregards the possibility that subjects change their decision criteria in the course of the experiment. It furthermore assumes that all subjects mechanistically respond to the stimuli in the same way, so that it does not allow for the fact that different subjects may use different decision criteria.

An alternative approach recognizes the active role of the subject in the experimental task:

> Most recent approaches to the study of sensation and perception (...) have recognized the active role the subject takes when he tries to make sense of the experimental demands and the properties of the stimulus array. This active attempt at "making sense" of the experimental situation is inescapable, forced in this situation as it is in his daily experience of his normal world (Baker et al., 1988: 402).

The response coincidence technique that is used in the analysis of the data explicitly takes into account the active status of the subject and is to be described as "a technique which examines how subjects categorize stimuli, as opposed to the more traditional techniques of analyzing how stimuli affect subjects" (Baker et al., 1988: 403).

The procedure of data analysis described here consists of two steps. In the first step, a response coincidence analysis is carried out. This is followed by a subject cluster analysis. The objective of this procedure is to group subjects according to their perception of the items, the underlying assumption being that each group consists of subjects who use similar decision criteria in the discrimination task. Both procedures are described below.

2.4.2. Response Coincidence Analysis.

The aim of response coincidence analysis is to compare the overall discrimination pattern of all participating subjects and to convert the obtained discrimination judgements into a dissimilarity metric. From this, a coincidence matrix for all subjects is constructed. The construction of this coincidence matrix will be briefly discussed on the basis of a hypothetical example.

Assume that 4 subjects took part in a discrimination experiment consisting of 5 items and that they were asked whether they could hear an intonational difference between the stimuli of each item. Subjects' judgements were recorded in terms of the labels 'same' (S) or 'different' (D). The hypothetical responses can be presented in the form of an incidence matrix in table 4.5:

	Item 1	Item 2	Item 3	Item 4	Item 5
Subject 1	S	S	D	S	D
Subject 2	D	D	S	S	S
Subject 3	D	S	D	S	D
Subject 4	D	D	D	D	D

Table 4.5: Incidence matrix of subject responses to 5 discrimination items in terms of the labels 'same' (S) and 'different' (D).

Each row in the incidence matrix displays the discrimination of the items of a subject and constitutes an 'incidence' response vector. Subsequently, a 'coincidence response matrix' is constructed from each response vector, which contains information about the pairwise treatment of all possible combinations of the items within each subject. This matrix is derived by taking the direct product of the response vector with itself for every subject. For subject 1 of this sample, the response vector is [S S D S D]. It's direct product is given in table 4.6:

$$\begin{bmatrix} S \\ S \\ D \\ S \\ D \end{bmatrix} [S\ S\ D\ S\ D] = \begin{bmatrix} SS & SS & SD & SS & SD \\ SS & SS & SD & SS & SD \\ DS & DS & DD & DS & DD \\ SS & SS & SD & SS & SD \\ DS & DS & DD & DS & DD \end{bmatrix}$$

Table 4.6: The derivation of the direct product of the response vector of subject 1.

Each cell in the resulting matrix is subsequently recoded in such a way that any pair of items which is discriminated in the same fashion is given the value of that discrimination. Any pair which is not discriminated in the same way is assigned an X. In this example, 'SS' is recoded as 'S', 'DD' as 'D' and both 'SD' and 'DS' as 'X'. The result of this conversion yields the matrix in table 4.7:

$$\begin{bmatrix} S & S & X & S & X \\ S & S & X & S & X \\ X & X & D & X & D \\ S & S & X & S & X \\ X & X & D & X & D \end{bmatrix}$$

Table 4.7: Recoded matrix for subject 1.

It can be observed that the matrix in table 4.7 is symmetrical along the main diagonal, with the result that only the information below the main diagonal is relevant, whereas that above the diagonal is redundant (or vice versa). Hence, the matrix can be reduced to table 4.8:

Subject 1

Item	1	2	3	4
2	S			
3	X	X		
4	S	S	X	
5	X	X	D	X

Subject 2

Item	1	2	3	4
2	D			
3	X	X		
4	X	X	S	
5	X	X	S	S

Table 4.8: Coincidence response matrices for subjects 1 and 2.

The matrix in table 4.8 gives an indication of the extent to which the various items are pooled together with other items. Matrix elements which contain the labels S and D indicate that the items were perceived in a similar fashion as either 'same' or 'different'. An X indicates that the two items were perceived in a dissimilar fashion. After the coincidence matrices for all subjects have been constructed, a dissimilarity measure is calculated between all pairs of subjects. To obtain this dissimilarity measure for each pair of subjects, the coincidence matrices are compared cell by cell. The number of cell mismatches between the matrices is calculated and divided by the number of item pairs. This yields a dissimilarity metric which ranges from zero for perfect correspondence to one for total lack of correspondence. In order to calculate the dissimilarity metric for subjects 1 and 2, the coincidence matrices for these subjects in table 4.8 are compared cell by cell. Out of 10 possible item pairings, six pairs were assigned to different categories which results in a dissimilarity metric of 0.60. If this procedure is applied to all the subjects of the present example, the intersubject dissimilarity matrix in table 4.9 is obtained:

	Subj.1	Subj. 2	Subj. 3	Subj. 4
Subj. 1	-			
Subj. 2	0.60	-		
Subj. 3	0.40	0.70	-	
Subj. 4	0.90	0.90	0.70	-

Table 4.9: Intersubject dissimilarity matrix for our hypothetical data.

From this matrix, it can be seen that subjects 1 and 3 are the most similar in their response pattern, since they have the lowest dissimilarity metric (0.40). The responses of subjects 1 and 4 and subjects 2 and 4 are most dissimilar (0.90), which correlates with a big difference in their overall discrimination. This dissimilarity metric in the intersubject dissimilarity matrix can be shown to be equivalent to a squared distance in the Euclidean sense (Clifford & Stephenson, 1975: 66, cited in Baker et al.), with the result that the matrix can be used as an input to Ward's method of hierarchical cluster analysis.

2.4.3. Cluster Analysis: The Search for Subject Groups.

In the next stage, the dissimilarity matrix for all pairs of subjects is used as an input to hierarchical cluster analysis (SPSS-X, 1988). This technique allows the experimenter to establish groups of subjects which can be assumed to have similar response patterns. The underlying assumption is that the subjects within each cluster use a similar discrimination strategy.

The aim of applying response coincidence analysis to the data in this experiment is to establish whether it is possible and necessary to distinguish groups of subjects who differ in their perceptual response to the discrimination items of this experiment. The main variable of interest is the distribution of the number of 'different' judgements along the rise-alignment continuum. With regard to this variable, it was concluded that there are no significant discrimination peaks in the average discrimination curve based on the results of all the 94 participating informants. It remains nevertheless conceivable that a discrimination function based on the results of a subgroup of subjects does have a discrimination peak, whereas this of other subject groups does not. The averaging that has taken place to arrive at the discrimination function for the data pooled across all subjects may have obscured such a peak. In this section, we show that response coincidence analysis, coupled to cluster analysis, can be regarded as a valid tool to reveal groups of subjects who do or do not show discrimination peaks. For this purpose, the technique will be applied to a set of hypothetical data.

Let us take a situation in which 10 subjects are presented with 6 discrimination items taken from a pitch movement alignment continuum, and that they were required to judge these stimuli in terms of the labels 'same' or 'different. The hypothetical data might look as follows: the responses of 5 subjects are identical in that they judged all the items to be 'same' (subject group 1). The 5 other subjects responded 'same' to items 1, 2, 3, 5 and 6. All these subjects judged item 4 to be 'different' (subject group 2). The coincidence matrix for all the items as judged by group 1 is given in table 4.10:

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
ITEM1	S					
ITEM2	S	S				
ITEM3	S	S	S			
ITEM4	S	S	S	S		
ITEM5	S	S	S	S	S	
ITEM6	S	S	S	S	S	S

Table 4.10: Coincidence matrix the items as judged by the subjects in group 1 who label all 6 discrimination items 'same'.

The dissimilarity metric for each subject in this group will other subjects in this group is 0.00, i.e. there is a perfect correspondence in the way these subjects respond to the individual stimuli.

The coincidence matrix for the items as judged by the subjects in the second group is given in table 4.11:

	ITEM1	ITEM2	ITEM3	ITEM4	ITEM5	ITEM6
ITEM1	S					
ITEM2	S	S				
ITEM3	S	S	S			
ITEM4	X	X	X	D		
ITEM5	S	S	S	X	S	
ITEM6	S	S	S	X	S	S

Table 4.11: Coincidence matrix for all items as judged by subjects in group 2, who label all the discrimination items with 'same', except for item 4, which is labelled as 'different'

The dissimilarity metric for each subject in this group is 0.00 as well, since the discrimination results of all the subjects are in perfect agreement. The metric between the subjects of this group and those of the first group amounts to 0.285 for each subject pairing. This yields the following subject coincidence matrix in table 4.12:

	SUB1	SUB2	SUB3	SUB4	SUB5	SUB6	SUB7	SUB8	SUB9	SUB10
SUB1	0.00									
SUB2	0.00	0.00								
SUB3	0.00	0.00	0.00							
SUB4	0.00	0.00	0.00	0.00						
SUB5	0.00	0.00	0.00	0.00	0.00					
SUB6	0.28	0.28	0.28	0.28	0.28	0.00				
SUB7	0.28	0.28	0.28	0.28	0.28	0.00	0.00			
SUB8	0.28	0.28	0.28	0.28	0.28	0.00	0.00	0.00		
SUB9	0.28	0.28	0.28	0.28	0.28	0.00	0.00	0.00	0.00	
SUB10	0.28	0.28	0.28	0.28	0.28	0.00	0.00	0.00	0.00	0.00

Table 4.12: Subject coincidence response matrix expressing the dissimilarity in reponse pattern to the discrimination items.

When submitted to hierarchical cluster analysis, the following subject grouping is obtained :

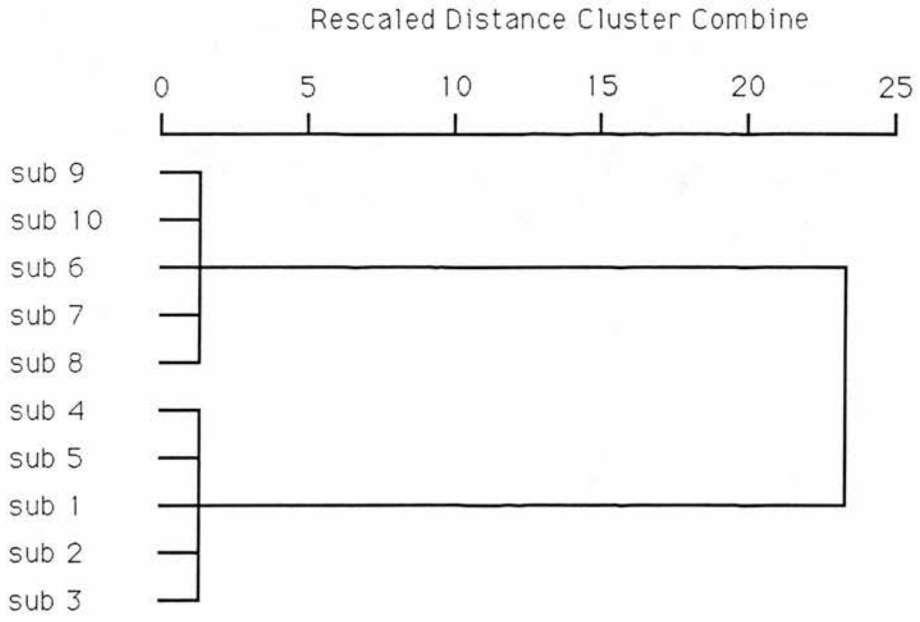


Table 4.13: Dendrogram based on a cluster analysis of the response coincidence matrix in table 4.12.

Although the average discrimination function based on the results of all subjects in this hypothetical experiment has a clear discrimination peak at item number 4, the reality is that this peak is a feature of one group of subjects only, while the other group does not discriminate between any of the items. In this situation, the coincidence analysis proposes a grouping of subjects based on similarities in their overall response to all the items in the experiment. Drawing up a separate discrimination curve for each group reveals the difference.

It is clear that the hypothetical data presented here provide a simple case, whereas the data of this experiment are much more complex, and therefore it is difficult to provide 100% certainty that the technique will handle these complex data in a similar way. Two additional arguments need to be considered.

In the first instance, the technique was validated by Baker et al. (1988) on the basis of tone identification data by native speakers of Chinese reported in Connell et al. (1983). In this experiment, 28 informants were presented with three syllables carrying 4 Mandarin tones, which were synthetically altered in order to simulate intonational changes in these syllables in sentence-final position. Subjects were required to label these experimental stimuli. Application of response coincidence analysis and cluster analysis yielded four groups of subjects, which have quite different identification functions in that the overall level of identification differs significantly between the groups. Moreover, small but consistent differences in the location of the tonal category boundaries became apparent as well as one case where one group did not have a category boundary at all for one specific tone. This clearly suggests that response coincidence analysis has potential for revealing consistent discrimination differences between subjects in this experiment.

The second argument in favour of applying response coincidence analysis is that it is an automated process in that the data are handled by computer. It would be impossible to process all these data manually by comparing the discrimination curves for each subject with the curves of the 93 other participants in the experiment and provide formal criteria to separate out groups of subjects.

2.4.4. Subject Groups.

The construction of the response coincidence matrices was achieved by a computer programme especially written for this purpose. The obtained subject matrices were subsequently analysed by a conventional hierarchical clustering technique using the SPSS-X-software package (1988). Clusters were based on Wards method. The response coincidence matrices are given in the appendix (tables A11-A14). The dendrograms for each of the cluster analyses are given in figures 4.18 to 4.21 below:

```
                       Rescaled Distance Cluster Combine

  C A S E      0         5        10        15        20        25
Label     Seq  +---------+---------+---------+---------+---------+

SUB11      11   -+-+
SUB17      17   -+ +-+
SUB10      10   ---+ +-----+
SUB12      12   -----+     +---+
SUB15      15   -----------+   +---------------------+
SUB9        9   ---------------+                     +-----------+
SUB5        5   -------+---------+                   |           |
SUB7        7   -------+         +-------------------+           |
SUB8        8   -----------------+                               |
SUB4        4   -----------------------+-----------+             |
SUB6        6   -----------------------+           +---------+   |
SUB1        1   -------------------------------+---+         |   |
SUB13      13   -------------------------------+             +---+
SUB2        2   ---------+---------+                         |
SUB18      18   ---------+         +---------------------+   |
SUB3        3   -------------------+                     +---+
SUB14      14   -------------------------+---+           |
SUB19      19   -------------------------+   +-----------+
SUB16      16   -----------------------------+
```

Figure 4.18: Dendrogram for subjects in the 50-70 msec AB experiment.

```
SUB10      10   -+
SUB11      11   -+-----------------------------+
SUB15      15   -+                             +-------+
SUB19      19   -----------+-------------------+       |
SUB22      22   -----------+                           |
SUB13      13   -------+-----+                         +---------+
SUB17      17   -------+     +-------------------+     |         |
SUB20      20   -------------+                   |     |         |
SUB6        6   -----+---+                       +-----+         |
SUB18      18   -----+   +---------+             |               |
SUB12      12   ---------+         +---+         |               |
SUB14      14   -------------------+   +---------+               |
SUB16      16   -----------------------+                         |
SUB2        2   ---+-----------------+                           |
SUB3        3   ---+                 +-------------------+       |
SUB1        1   ---------------+-----+                   |       |
SUB8        8   ---------------+                         +-------+
SUB4        4   -------------------------+---------+     |
SUB5        5   -------------------------+         +-----+
SUB7        7   -----------------+-----------+     |
SUB21      21   -----------------+           +-----+
SUB9        9   -----------------------------+
```

Figure 4.19: Dendrogram for subjects in the 50-70 msec BA experiment.

```
                 Rescaled Distance Cluster Combine

  C A S E      0         5        10        15        20        25
Label     Seq  +---------+---------+---------+---------+---------+

SUB2        2   -+
SUB25      25   -+
SUB20      20   -+-+
SUB1        1   -+ +-+
SUB24      24   ---+ +-+
SUB12      12   -----+ +-------+
SUB23      23   -----+ |       +---------------------------+
SUB8        8   -------+       |                           |
SUB9        9   ---------------+                           |
SUB3        3   -------------+-----------+                 +-----+
SUB22      22   -------------+           +-------------+   |     |
SUB16      16   -------------------------+             |   |     |
SUB6        6   -----------------+-------------+       +---+     |
SUB13      13   -----------------+             |       |         |
SUB10      10   ---------+-+                   +-------+         |
SUB19      19   ---------+ +---------+         |                 |
SUB14      14   -----------+         +-+       |                 |
SUB21      21   ---------------------+ +-------+                 |
SUB18      18   -----------------------+                         |
SUB4        4   ---------------------------+---------+           |
SUB11      11   ---------------------------+         |           |
SUB7        7   -------------------+-------------+   +-----------+
SUB15      15   -------------------+             +-+ |
SUB5        5   ---------------------------------+ +-+
SUB17      17   -----------------------------------+
```

Figure 4.20: Dendrogram for subjects in the 90-110 msec AB experiment.

```
                         Rescaled Distance Cluster Combine

 C A S E          0         5        10        15        20        25
Label        Seq  +---------+---------+---------+---------+---------+

SUB18         18   -+
SUB27         27   -+
SUB20         20   -+---+
SUB14         14   -+   +---------------------------------------+
SUB17         17   ---+-+                                       |
SUB21         21   ---+                                         +---+
SUB7           7   -------+-+                                   |   |
SUB19         19   -------+ +---------+                         |   |
SUB28         28   ---------+         +-------------------------+   |
SUB22         22   ---------+-----+   |                             |
SUB30         30   ---------+     +---+                             |
SUB2           2   -----+-------+ |                                 |
SUB16         16   -----+       +-+                                 |
SUB1           1   -------------+                                   |
SUB4           4   -------------------------+-----------------+     |
SUB11         11   -------------------------+                 |     |
SUB8           8   -----------+-----+                         |     |
SUB23         23   -----------+     +---+                     +-----+
SUB9           9   -----------------+   +-----------------+   |
SUB15         15   ---------------------+                 |   |
SUB3           3   -----------------+-----------+         |   |
SUB29         29   -----------------+           +---+     +---+
SUB5           5   -----------------------------+   +---+ |
SUB13         13   ---------------------------------+   | |
SUB6           6   -----------------------------+-----+ +-+
SUB12         12   -----------------------------+     | |
SUB25         25   -----------------------+---+       +-+
SUB26         26   -----------------------+   +---+   |
SUB24         24   ---------------------------+   +---+
SUB10         10   -------------------------------+
```

Figure 4.21: Dendrogram for subjects in the 90-110 msec BA experiment.

In establishing groups of subjects, the problem of the statistical decision criterion arises: i.e at which clustering level are the clusters heterogeneous enough to justify separation of subjects into different groups. There are no general criteria available for this purpose (Seber, 1984) since the distributional properties of distance measures do not lend themselves to the 'normal' distribution assumptions. The ad hoc solution proposed by Baker et al. (1988), consisting of analysing randomized discrimination data does not seem to work for these results. Consequently, the highest clustering level has been used as an indication of the final groupings of subjects.

For both groups of subjects established in this way, average discrimination for each of the stimulus conditions is given in figures 4.22- 4.25:

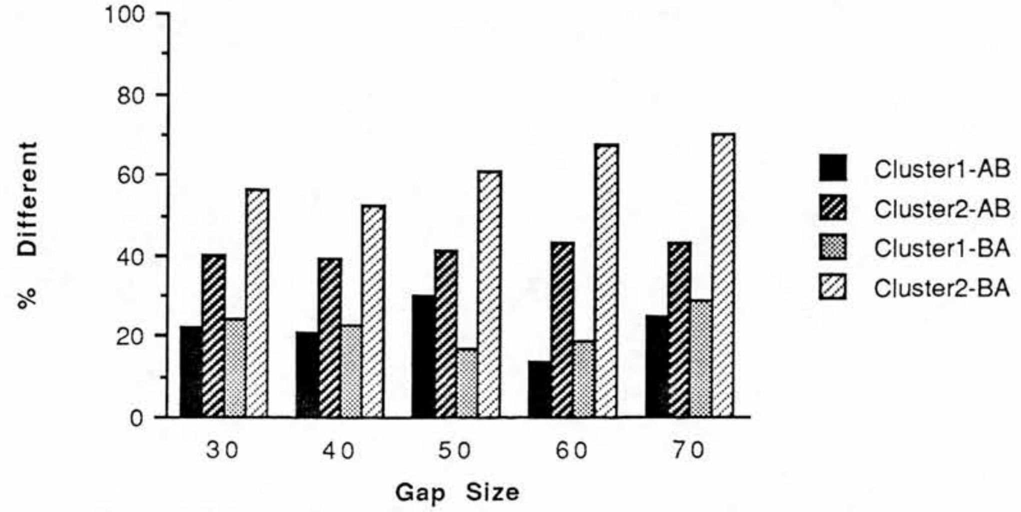


Figure 4.22: Average discrimination in items with slope 50 msec as a function of gap size for the clusters identified by the cluster analysis in AB and BA presentation order.

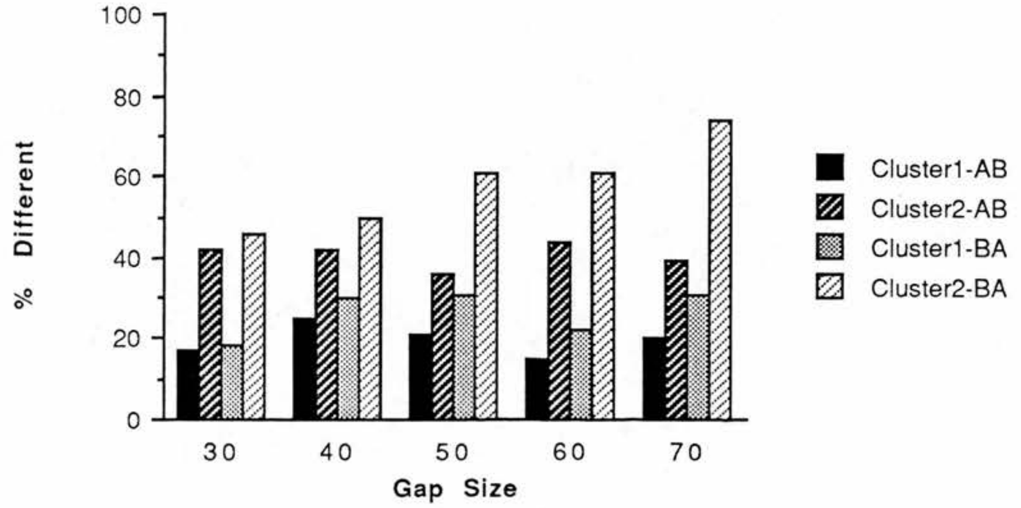


Figure 4.23: Average discrimination in items with slope 70 msec as a function of gap size for the clusters identified by the cluster analysis in AB and BA presentation order.

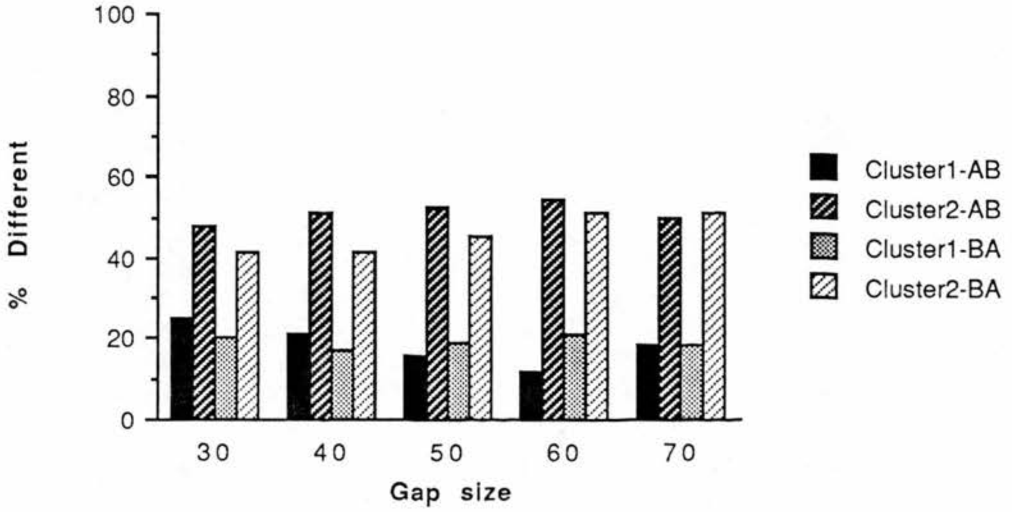


Figure 4.24: Average discrimination in items with slope 90 msec as a function of gap size for the clusters identified by the cluster analysis in AB and BA presentation order.

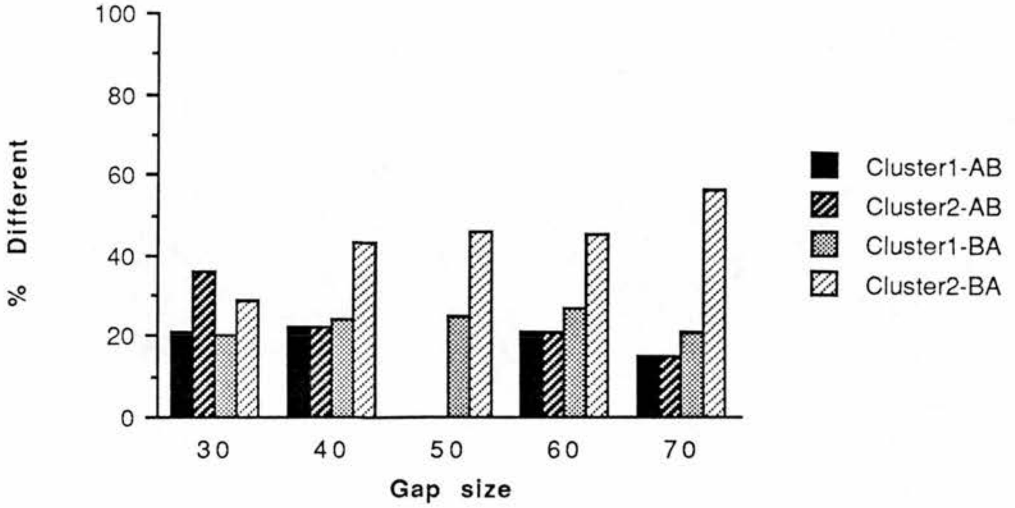


Figure 4.25: Average discrimination in items with slope 110 msec as a function of gap size for the clusters identified by the cluster analysis in AB and BA presentation order. Observations for gap 50 msec AB are missing due to technical problems.

From figures 4.22 - 4.25, it can be seen quite clearly that the subjects are clustered as to how well they generally discriminate the stimuli in the items. Cluster 1 is constituted by subjects who have great difficulty in discriminating the stimuli. Averaged over the total number of items, this group of subjects discriminates the items in 22% of the cases. The subjects in this group will be called *non-discriminators*. The subjects in cluster 2 generally discriminate the stimuli much better: they judge half of the presented stimuli (49,9%) as different. They will be referred to as *discriminators*. A paired t-test on the proportions of different judgements in the discriminator and non-discriminator groups shows that this overall difference in discrimination is significant at $p<0.001$.

From these results, it can be concluded that the subjects who have taken part in the experiment, do not operate on the same discrimination criteria. It is not possible to specify these criteria explicitly, since there is no detailed information regarding the subjects themselves. The clusters, however, give a clear indication that one group of subjects discriminates markedly better than the other group.

The fact that there are non-discriminators is not very surprising. In 't Hart's experiment on the discrimination of pitch excursion size differences ('t Hart, 1981), 42% of the informants are not able to discriminate. It is however remarkable to observe that in this experiment more than half of the informants belong to the non-discriminator group (60%). There may be a variety of reasons why the subjects in this group do not discriminate the stimuli in the items. It is possible that they are not really interested in the task and therefore simply perform badly. More importantly, it may be indicative of the fact that the acoustic differences in alignment used in our discrimination experiment are not at all relevant to the perception of the contours under investigation, which corroborates the fact that there are no discrimination peaks in the alignment continuum.

Since the subjects who have taken part in this experiment can be shown to use different discrimination criteria, it was decided to split up the average discrimination function in figure 4.15 into a separate one for the discriminator and non-discriminator groups. This has to be considered as an essential further step in the analysis, since it may reveal additional differences in discrimination strategies, which potentially relate to the perception of pitch movement alignment. Any consistent differences are likely to be reflected in the groupings of subjects proposed here, but they cannot be derived from figures 4.22 - 4.25, since these refer to average overall discrimination of the items.

The discrimination functions for the discriminator and non-discriminator group are given in figure 4.26:

Figure 4.26: Average discrimination as a function of mean rise position (defined in terms of rise endpoint) for the discriminator (D) and non-discriminator (ND) group.

Apart from the overall difference in discrimination between the two groups, figure 4.26 does not reveal any marked differences regarding the perception of pitch movement alignment: the discrimination functions for both groups of subjects are essentially identical in that there are no significant discrimination peaks in the continuum.

Like the parameter of pitch movement alignment, the OE was analyzed separately for the two groups of subjects. This is summarized in figure 4.27:

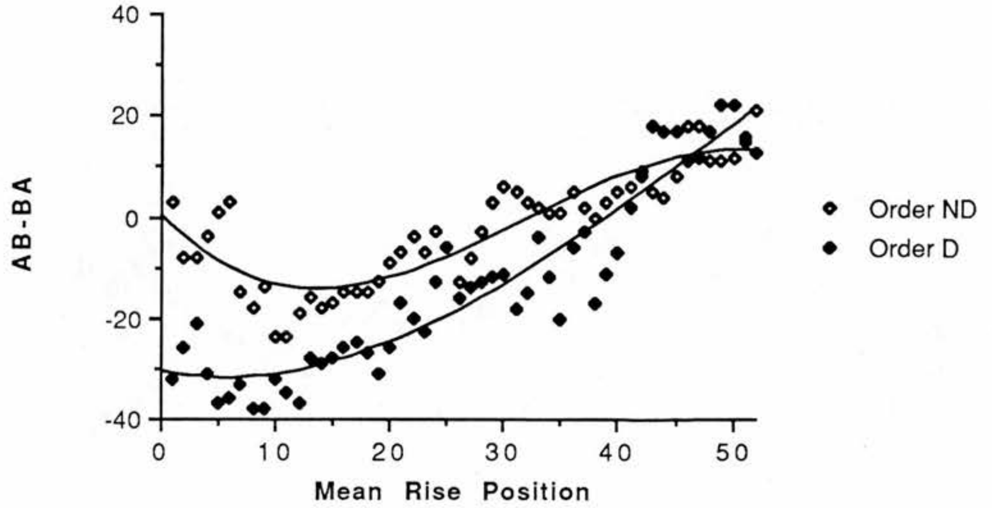


Figure 4.27: Quantification of the order effect (AB-BA) as a function of mean rise position for the discriminator (D) and non-discriminator (ND) group. Total number of observations is 20,840.

A paired t-test on the quantifications of the OE in the discriminator and non-discriminator group indicates a statistically significant difference in discrimination at p<0.001. Figure 4.27 shows that the OE is more outspoken in the discriminator group.

3. Discussion.

3.1. Pitch Movement Alignment.

The data obtained in this experiment provide a rather complicated picture of rise discrimination in the Dutch hat pattern. Regarding the variable ORDER, it can be observed that the BA presentation order is significantly better discriminated than the AB order. This means that items, in which the stimulus with the later rise location is presented first, are significantly easier to discriminate. It should be pointed out, however, that this trend is not uniform but is clearly related to the location in the stimulus continuum. In the left-hand side of the continuum, the BA is better discriminated than the AB order. In the central area of the continuum, OE disappears and is reversed in the right-hand side of the stimulus array.

The physical dimensions of GAP and SLOPE have been found significant in the BA presentation order only and their effects on discrimination are qualitatively different. Increasing the gap size leads to an improvement of discrimination, whereas an increase in the slope of pitch movements causes a deterioration of discrimination performance.

The most important conclusion of this experiment relates to the variable rise-alignment in the experimental syllable. Concerning this variable, the analysis of variance indicated a signficant interaction with presentation order of the stimuli which was as follows: in the AB presentation order, delaying the rise in the experimental syllable yields a small increase in discrimination. In BA presentation order, however, delaying the rise causes the alignment differences to be discriminated progressively worse. It was also found that there are no significant interactions between ALIGNMENT x GAP and ALIGNMENT x SLOPE, so that it can be concluded that alignment differences are discriminated in a similar fashion in all gap and slope conditions. A re-analysis of the data in terms of the 'mean rise position' of the individual discrimination items reveals an almost entirely flat discrimination curve, with no significant discrimination peaks. Furthermore, it is observed that this flatness is characteristic for the curves based on MRP defined in terms of both rise onset and rise endpoint. It is to be emphasized that the absence of discrimination peaks is unlikely to have been caused by the averaging procedures in processing the data, since this technique has been tested by Hill & Reid (1977). They have shown that consistent discrimination peaks in the raw data are reflected by this averaging process. The results of the identification experiment are entirely consistent with the absence of peaks in the discrimination function, in that the identification function reveals a very gradual change in the labelling judgements from 'high' in the early rise positions to 'low' in late rise locations.

As far as the overall height of the average discrimination function is concerned, it has to be observed that average discrimination of the items is extremely low, i.e. 34%. This suggests that informants generally do not discriminate between the stimuli in the items. This conclusion is reinforced by the results of a response co-incidence analysis, which indicate that two groups of subjects have to be distinguished. The majority of subjects (60%) are classified as non-discriminators in that their average discrimination is only 22%. The remaining subjects belong to the discriminator group, in which average item discrimination amounts to 49,5%. Taken together, these data clearly suggest that informants do not reliably discriminate between the stimuli taken from the alignment continuum.

The absence of discrimination peaks and the smooth transition in the labelling judgements can be taken as evidence that the alignment continuum is not perceived in terms of categories. On this basis, it can be concluded that all the realizations of the hat pattern in the stimulus continuum relate to a single perceptual category, although there are clear physical differences in rise-alignment.

At a very fundamental level of analysis, the absence of categorical boundaries in the stimulus continuum indicates that the precise location of the rises in the first accented syllable of the flat hat intonation pattern is irrelevant to the categorization of this pattern. As a result, it is argued that rise alignment per se cannot constitute the perceptual essence of of the flat hat. If it were, there should be an indication of this in the form of a differential categorization of the pitch contours in the stimulus continuum. This is clearly not the case.

The phonetic relevance of pitch movement alignment to the specification of the flat hat pattern has always been implicit in the grammar of Dutch intonation, in that an early rise position is regarded as one of its essential characteristics. From Collier (1972), it appears that from a purely phonetic point of view, the vast majority of rises in the flat hat in the IPO corpus are found to be of type 1. Despite this phonetic systematicity, it is clear that the perceptual evidence provided in this experiment suggests that a more general characterization of the hat pattern is required. We suggest tat the essence of the hat pattern at an abstract level of analysis is captured by a prominence-lending rise, followed by a prominence-lending fall. The precise location of this prominence-lending rise in the syllable is allowed to vary a great deal, since these differences are not perceived by informants. Depending on the number of accents in the utterance, the pattern can take different shapes. If there is only one accent to be realized, both rise and fall occur on the same syllable. If there are two accented syllables in the utterance the first accent is realized by the rise, whereas the second accent is realized by the fall. Between both movements, declination remains at a high level.

The fact that the location of the rise in the accented syllable is not decisive to the categorization of the hat pattern indicates that informants do not listen atomistically, but rather that the contour is perceived wholistically as an instance of the hat pattern on the basis of its general form. The relatively large differences in rise alignment are not perceived, although the informants were specifically instructed to listen to these differences only. As such, a wholistic component in the perception of contours is to be recognised.

The evidence of this experiment also indicates that the phonetic identity of the pitch movements is not exclusively determined by the intonation pattern in which they occur, a principle which was explicitly formulated in 't Hart & Collier (1979):

> The nature and the order of all the pitch movements in an utterance are determined by the intonation pattern (p. 399).

By doing this, the IPO analysis does not take into account the possibility that there may be other factors involved in determining the concrete phonetic realization of the pitch movements in a given pattern. These factors can be assumed to be of a purely phonetic nature, as shown by Silverman & Pierrehumbert (in press), who indicate that vowel duration, speech rate and prosodic context are relevant in this perspective. In the IPO approach, intonation is reduced to the study of pitch phenomena, and does not take into account these factors.

It should also be emphasized that the results presented here invalidate the uniformity hypothesis. Although informants seem to be able to distinguish three locations of pitch rises in an utterance final syllable (Collier, 1975a), they do not discriminate alignment differences in the hat pattern. This indicates that the tolerance levels established in an isolated context cannot be extrapolated uniformly to other prosodic environments, since this may lead to incorrect observations regarding the categorization of pitch movements.

As a final point in this discussion, it is worth pointing out that the results presented here do not corroborate those reported in Verhoeven (1987). In this experiment, the discrimination of a similar rise alignment continuum was investigated using a 2IAX paradigm. At the time, the results of this experiment were interpreted as an indication of categorical perception, since a significant discrimination peak was found. In retrospect it is noted that this peak is located at the acoustic boundary between two segments in the experimental syllable. This is to say that for both gap sizes, subjects discriminated between pitch movements in an item, if the first movement was entirely situated within the vowel nucleus of the experimental syllable, whereas the endpoint of the second rise was located in the following consonant with 30 and 40 msec respectively.

The synthesis method used in preparing the stimuli in both experiments may hold the key to explaining the incompatibility of the results. In Verhoeven (1987), the stimuli were produced by means of the Holmes synthesis-by-rule system, in which the sounds are stored as steady-state spectral values. In the synthesis stage, these values are interpolated by means of a simple transition algorithm. Consequently, pseudo-acoustic boundaries are introduced in the synthesized speech signal. Since the discrimination peak coincides with these pseudo-acoustic boundaries, it may be suggested that its presence may be an artefact of segmental boundaries, a situation which cannot have occurred here, since the stimuli were derived from natural speech.

3.2. The Order Effect.

In processing the results, presentation order has been singled out as an experimental variable and a highly systematic OE is observed. This effect is seen to depend on the location of the pitch movements in the experimental continuum, in the sense that it was largest at the extremes of the continuum. In the middle of the continuum, it was virtually absent. Although presentation order was originally intended to be a control variable, it is appropriate to discuss its effect in some detail here.

The term order effect refers to the phenomenon in which the order of successively presented stimuli in discrimination experiments affects the discrimination ability of informants. This effect is well-established in psychological research on a wide variety of modalities, most particularly concerning the perception of duration, in which the effect is referred to as the time-order error. Two kinds of errors are traditionally distinguished, i.e. a positive and a negative one. In the former, the error is one of overestimation of the standard stimulus, whereas the latter is related to an underestimation of the standard (Woodrow, 1951). As an explanation of this phenomenon, several proposals have been put forward. In earlier work, the effect was assumed to result from memory or perceptual processes. More recently, OE is considered to result from subjects' response biases. In discrimination experiments on speech, OE has seldom been investigated systematically and has largely passed unnoticed due to the averaging procedures that have been used for obtaining the discrimination functions. Some notable exceptions deserve attention here.

Repp et al. (1979) investigated categorical perception in a synthetic continuum of isolated vowels (i, ɪ, ɛ), in which auditory memory was degraded by interference and time delay. They found that:

> The majority of the stimulus pairs received more 'different' responses when the first stimulus in a pair had a lower [sic] position on the continuum (i.e. was more [i]-like) than the second stimulus. (Repp et al., 1979: 138).

They further observe that "this effect disappeared or was even reversed at the right end of the continuum [i.e. when the sound was most ɛ-like]" (p. 138). To account for this phenomenon, Repp et al. refer to the negative time order error. This causes an underestimation of vowel duration in the first stimulus, so that, besides acoustic cues, there may be durational cues facilitating discrimination of one order of presentation over the other.

Uselding (1977) reports on OE in discrimination experiments on a Voice Onset Time (VOT) continuum. These results suggest that items are discriminated best if the first stimulus is a voiceless one. His data however also reveal a reversal of the order effect at the right of the continuum (in the area with voicing lag): in this area, the items in which the first stimulus had the smallest voicing lag were discriminated better than the reverse order of presentation. Similarly, in stimuli with negative VOT, better discrimination is suggested if the VOT value of the first stimulus is smallest.

More recently, significant order effects have been reported in intonation studies on the perception of F0 excursion size differences (Batliner & Schiefer, 1987) and peak alignment (Kohler, 1987). In the former, it is concluded that OE is not a purely psychophysical phenomenon, since it occurs in both speech and non-speech stimuli. The latter accounts for the OE in terms of the specific characteristics of the experimental paradigm.

It may be inappropriate to compare the results of these experiments on segmental and suprasegmental aspects of speech directly, but a noticeable trend in all the data is that OE is not uniformly present along the whole of the continuum under investigation. This is to say that in certain areas, the order effect manifests itself in a particular way, whereas in other areas it is reversed or even disappears completely. Therefore, a description of OE in terms of "the AB order is better discriminated than the BA order" (Batliner & Schiefer, 1987) has to be regarded as an oversimplification, if it is intended to mean that the AB order is always better discriminated than the BA order in a given experiment.

The results obtained in the experiments reported here confirm the observations in other experiments in which OE was controlled. The OE here is very outspoken and highly systematic. In the following section, the OE is accounted for in terms of phonetic prototypes.

In the original account of the order effect reversal (Verhoeven, 1988), it was argued that this effect cannot be accounted for in terms of a positive or negative time-error, since this would suggest that durational cues contribute to a better discrimination in one order over the other. Research has indeed shown that informants are quite insensitive to fairly large differences in rise durations (Collier, 1983). This conclusion also emerges from this experiment, since the variable of rise duration (slope) is shown to be only marginally relevant to discrimination. As an alternative to the time-order error, the prototype hypothesis is proposed, which is based on studies on phonetic prototypes (Samuel 1977, 1982).

In Samuel (1977), a categorical perception experiment is reported, in which informants were trained to discriminated VOT differences. The results suggest that discrimination performance improved at the category border areas, whereas it did not improve near the category centres (the prototypes). Samuel concluded that:

> with training, items further from the prototype would be discriminable from the prototype, and thus discriminable from each other (Samuel, 1977: 307).

If this finding is applied to our data, a coherent account of the order effect becomes possible. The discrimination results suggest that all the rises in the continuum are taken from a single alignment category, which can be assumed to have a prototypical realization. Thus, it is reasonable to assume that informants are unable to tell the difference between two rises, provided that they are both realizations of the prototype, irrespective of the presentation order of the stimuli. This can account for the fact that no OE is present in the central area of the alignment continuum, if we assume that the endpoint of the prototypical rises in this intonation pattern is located in the second half of the experimental syllable.

If informants are asked to compare non-prototypical rises in the discrimination items, one rise necessarily has to be a better example of the category (or more prototypical) than the other, and it is hypothesized that it is precisely this inequality that causes the OE to appear. This is to say that in the presentation order in which the first stimulus contains the rise that is the poorest example of the category, a shift towards a rise which is more prototypical may induce a positive recognition effect. If confronted with the poorest example first, the informant may be in doubt whether it still is a member of the category. But on hearing a better example, he recognizes the first one as being similar enough in order to belong to the same category. Hence discrimination is low if a temporal shift towards the prototype is involved.

If on the other hand the rise in the first stimulus is a better non-prototypical member of the category, informants may be more inclined to notice the difference. In the case of the better example of the rise, the informant may already be doubtful about its identity. This doubt is aggravated by presentation of a still poorer example of the category in the second stimulus. Consequently, informants tend to discriminate this order of presentation more accurately, since a temporal shift away from the prototype is involved.

This hypothesis not only accounts for the data in this experiment, but is also consistent with data on temporal quantization of speech ('t Hart, 1979b) and categorical perception data on peak-alignment differences in intonation (Kohler, 1987). 't Hart (1979b) carried out a discrimination experiment, in which natural speech stimuli were to be compared with their temporal quantizations, in order to determine the extent to which differences between original and quantized utterances can be perceived. He observed a significant order effect, which was such that the items in which the original utterance (prototype) was presented first were discriminated significantly better than the reverse order presentations.

Kohler's data (Kohler, 1987) on the categorization of intonational peaks suggests the existence of two alignment categories: an early and a late one. This is reflected in the discrimination data by a clear discrimination peak. As far as OE is concerned, there are three areas in the continuum in which the effect is completely absent: at the centre of each alignment category and at the category boundary. In both areas, discrimination is equally good (or poor) in both orders of presentation. In the peripheral areas of each category, the OE is such that the best discrimination is obtained if the order involves a shift away from the centre of the alignment category (the assumed prototype). Therefore, if our interpretation is correct, the dramatic differences in discrimination for both presentation orders near the category boundaries in Kohler's data, can be logically accounted for.

Although this hypothesis is highly consistent with all the data available, we have been unable to provide additional perceptual evidence. An attempt was made to investigate the prototype hypothesis with respect to vowel duration in German. In this experiment(4) three continua of vowel duration were designed on the basis of the wordpairs 'Stadt' vs. 'Staat', 'Straffen' vs. 'Strafen' and 'Stattlich' vs. 'Staatlich'. Within each continuum, the stimuli were combined in AX discrimination items in different orders of presentation and presented to native speakers of German for discrimination. Furthermore, all the stimuli were presented for identification and at the same time informants were asked to provide a typicality judgement of vowel length. Although the results of this experiment are interesting in their own right in that clear evidence is found for categorical perception, they do not confirm the ex-

---

(4) A more detailed account of this experiment is given in the Postscript section.

pected relationship between the prototypical length of the vowels and the presentation order of the stimuli.

4. Conclusions.

The major conclusion of this experiment relates to the perceptual relevance of rise-alignment to the characterization of the hat pattern in Dutch. All the data presented here provide no evidence that informants categorize the hat pattern on the basis of rise-alignment. More specifically, it is observed that the raw discrimination functions do not reveal any consistent discrimination peaks and the average discrimination curve is almost entirely flat. In addition, a response coincidence analysis indicates that the majority of subjects belongs to the non-discriminator group. The remaining subjects are labelled as discriminators. The identification data finally are observed to be compatible with absence of discrimination peaks, in that they are characterized by a very gradual change from "high" to "low" judgements. On the basis of these observations, it can be concluded that the precise location of the rise in the pattern is not perceptually relevant and consequently need not be regarded as an essential aspect in the characterization of the pattern. It also suggests that the uniformity hypothesis is to be rejected in that it is unjustified to extrapolate knowledge about informants' perceptual tolerances to isolated pitch events to all prosodic environments.

It is also observed that the physical variables of GAP SIZE and SLOPE of the rises are only marginally relevant to discrimination. However, a significant and highly systematic order effect is observed, which can be characterized as order effect reversal. The effect is accounted for in terms of phonetic prototypes.

CHAPTER 5

The Discrimination of Fall Alignment in the Dutch Hat Pattern.

0. Introduction.

The results obtained in the previous experiment provided no indication of the fact that flat hat contours are categorized on the basis of the alignment of rises to the penultimate accented syllable in utterances. As a result, it is possible to argue that rise-alignment differences are not relevant to the perceptual identity of this pattern. In this chapter, alignment differences in falls are investigated with respect to the same intonation pattern.

The problem regarding fall alignment in the hat pattern is very similar to the one described in the previous chapter: the IPO corpus contains contours which are acoustically different with respect to the location of the falls, but the perceptual relevance of these differences has never been investigated. This is clearly the case for the distinction between 1A and 1C contours, which differ from each other in the alignment of the fall in the pattern. Since these contours have been observed to occur in Dutch (Collier, 1972), it will be examined here whether this difference is regarded as perceptually relevant by native speakers of Dutch. The aim of this experiment is twofold. On the one hand, the discrimination of fall alignment differences is investigated in order to establish whether native speakers of Dutch distinguish between 1A and 1C pitch contours. If they consider them as linguistically distinct, it can be expected that the discrimination results will show a clear discrimination peak in the alignment continuum, which can be assumed to represent heightened sensitivity near the category boundary. If they do not consider the contours to be linguistically different, it is predicted that there will be no discrimination peak.

On the other hand, a more global aspect is investigated. It was indicated in the previous experiment that the perception of the rise-alignment continuum may have been influenced by the location of the fall in the hat pattern. Consequently, a hierarchical component in the perception of pitch contours may have to be recognized. This will be more systematically examined in this experiment, in that the location of the rise in the hat pattern is manipulated as a separate independent variable aside from the location of the fall in the pattern. This will enable conclusions to be drawn about any relationship between rise location and the discrimination of fall alignment.

It should be pointed out that the distinction between A and C in the hat pattern can only materialize in very specific circumstances, in that it can only be associated with an utterance-final syllable, since the C movement has to occur utterance-finally. The result of this is that the falls A and C do not only differ in alignment, but also in their excursion size. In the IPO model, the fall of type A is described as a 'final fall' which occurs rather late in the syllable ('t Hart & Collier, 1975). The fall of type C is described as follows :

> (...) during the last 20 to 50 ms of phonation in the utterance, F0 goes down rapidly to an immaterial value. This movement, although probably a mere relaxation phenomenon, is perceptually relevant: its omission in the stylized contour is readily noticed. Its position in the final syllable should be very late (...) ('t Hart & Collier, 1975: 241).

The definitions given above indicate that the two falls are not only physically different in terms of their alignment, but they also involve a difference in excursion size in that the movement A is a complete fall to the lower declination line, whereas the movement C occurs so late at the end of utterance phonation that it is not possible to realize the fall completely. In the rest of this chapter, the term fall-alignment comprises these two acoustic characteristics.

1. Experimental Design.

The experiment consists of a 2IAX discrimination task, in which subjects are asked to judge melodic differences between stimuli taken from a fall alignment continuum. The independent variables in the experimental design are GAP size, rise LOCATION and fall ALIGNMENT.

1.1. Stimuli.

The stimuli for this experiment are based on the utterance 'Renaat is naar Parijs' (Renaat is in Paris). This sentence was read 20 times by a male native speaker of Dutch, who realized the utterances with a flat hat pitch contour. His deliveries were recorded on tape with digital recording facilities and from these recordings, the most natural and fluent delivery was chosen as the basis for the stimuli. These were prepared by the method of modified resynthesis, which was described in detail in the previous chapter. It will suffice here to discuss the F0 manipulations for these stimuli.

The ILS analysis file resulting from the acoustic analysis was copied three times so that three segmentally identical utterances were obtained. In each analysis file, the original pitch contour was replaced by an artificial one. Before the F0 manipulations were implemented, the utterance was segmented in order to determine the F0 anchor points for the artificial pitch contour. The relevant segment boundaries of the utterance are given table 5.1:

Segment	Start	End	Duration
[ʀə]			
[n]	51	79	280
[ɑː]	79	96	170
[t]	96	104	80
[ɪs nɑːʀpɑ]			
[ʀ]	162	168	60
[ɛɪ]	168	190	220
[s]	190	214	240

Table 5.1: Segment boundaries of the syllables with which the rise and fall of the hat pattern are associated. Start and endpoints are given in ILS frame numbers and duration in milliseconds.

The contours that were implemented on the utterances differed mainly in the acoustic characteristics of the rise associated with the syllable 'naat' of 'ReNAAT'. In the first utterance, the rise constituted an unambiguous example of a type 1 movement: it covered an excursion of 4 semi-tones over 100 msec and its endpoint was located at 30 msec after vowel onset. The stimuli derived from this contour will be referred to as rise condition 1. The overall specification of this experimental contour is given in figure 5.1:

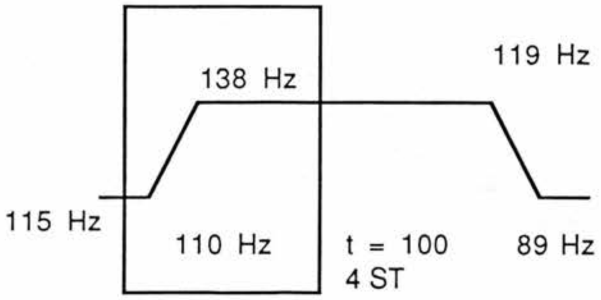


Figure 5.1: Specification of the contour with unambiguous rise 1. Duration of the rise is 100 msec and excursion size 4 semi-tones. Declination is omitted from the graph.

In the second utterance, the rise associated with the syllable 'naat' was an unambiguous example of a type 3 movement. It had an excursion size of 6 semi-tones and a duration of 150 msec. The endpoint of the rise coincided with the last frame of the vowel in 'naat'. This contour is detailed in figure 5.2 and the stimuli with this contour will be referred to as rise condition 3:

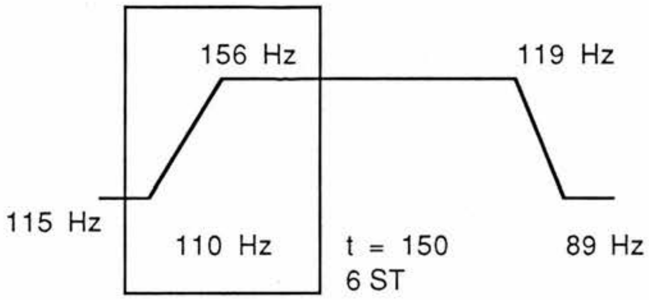


Figure 5.2: Specification of the contour with unambiguous rise 3. Duration of the rise is 150 msec and excursion size 6 semi-tones. Declination is omitted from the graph.

The pitch contour of the third utterance contained a rise which was acoustically ambiguous between rise 1 and rise 3. Its excursion size was 5 semi-tones with a duration of 120 msec. The endpoint was located at 70 msec after vowel onset, which is exactly half-way between the endpoints of the unambiguous rise 1 and rise 3. These stimuli will be referred to as the ambiguous rise condition. Their contour is illustrated in figure 5.3:

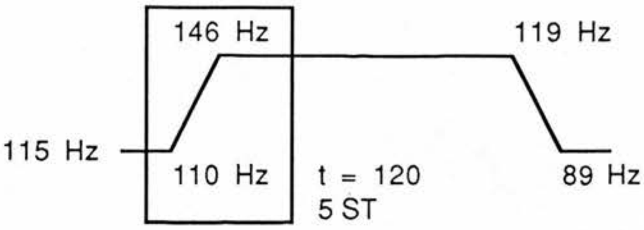


Figure 5.3: Specification of the contour with ambiguous rise. Duration of the rise is 120 msec and excursion size 5 semi-tones. The endpoint of the rise is located 70 msec after vowel onset. Declination is omitted from the graph.

In each of the utterances, the fall on the syllable 'rijs' of 'PaRIJS' was identical, i.e. it had a 5 semi-tone excursion and a duration of 80 msec. Its endpoint was located 80 msec after vowel onset. As a result of the unequal excursion size of the rises in the contours, the high declination line connecting the rise and the fall was slightly different in the three rise conditions.

From these three files, a stimulus continuum was derived by shifting the location of the fall to the right in steps of 10 msec. The effect of this is that the falling movement was gradually shifted into the postvocalic voiceless consonant. This procedure enabled the construction of an alignment continuum in which the earlier falls are fully realized, whereas the very late falls are incomplete, since they are partially situated inside the voiceless segment. By having the falls disappear in a voiceless segment, the manipulation of excursion size is achieved without altering the slope of the pitch movements as an additional variable. This results in a continuum in which some falls are more A-like in an IPO perspective, whereas others are more C-like. This is illustrated in figure 5.4 :

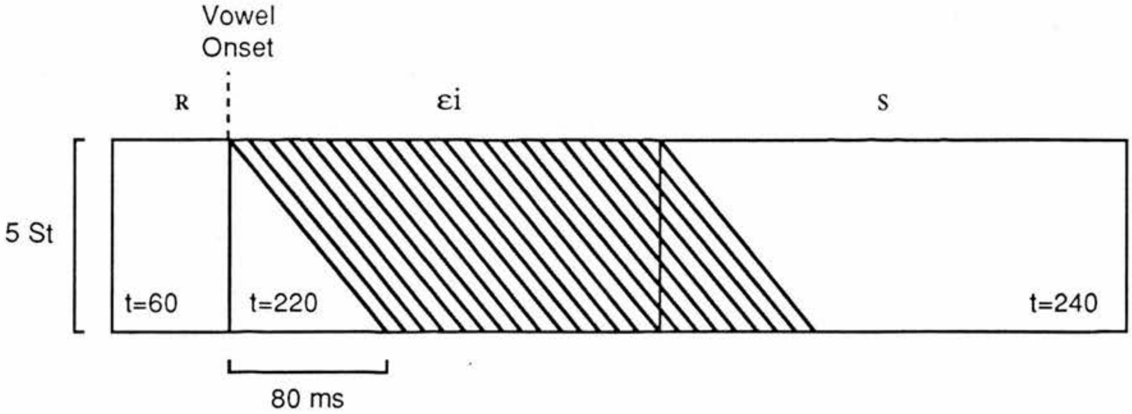


Figure 5.4: Illustration of the location of the falls in the alignment continuum.

The actual stimuli for the experiment were obtained by resynthesizing the analysis files by means of the ILS SNS algorithm. This yielded 69 stimuli (i.e. 23 x 3). The precise location of the falls in the experimental syllable of the stimuli is given in appendix A15.

1.2. Discrimination Items.

The stimuli in each continuum were paired into variable standard AX discrimination items with gap sizes 40 and 50 msec. This procedure yielded a total of 99 items. Due to the limited availability of the informants, it was decided not to control the presentation order of the stimuli in the items, nor to include any physically identical items. This made it possible to reduce listening time and to include multiple repetitions of the items: each item occurred 5 times in the test. This resulted in a total of 507 discrimination items.

1.3. Test Tapes.

The items were recorded on three different test tapes, each of which contained the items for a different rise condition. The items were grouped into two presentation blocks according to the gap size between the falls in the stimuli. The items with gap size 40 msec were presented in the first block, while those with gap 50 msec occurred in the second block. This was done in order not to predispose the informants towards the larger gap size. Within each block, the items were ordered randomly.
The inter-stimulus interval in each item was 500 msec and the inter-item interval amounted to 2,500 msec. There were no identifiers between the items, but after presentation of 10 discrimination items, the informants heard a short orientation signal. The total duration of each tape was approximately 30 minutes.

1.4. Informants.

Three groups of subjects were recruited from the student population of the 'Rijkshogeschool voor Vertalers en Tolken', Brussels (total N = 51). They were all native speakers of Dutch and took part on a voluntary basis. 11 subjects took part in the experiment for rise-condition 1, 20 in rise-condition 3 and 20 in the ambiguous rise condition. No subject participated in more than one experimental session.

1.5. Procedure.

For each of the experimental sessions, the informants were seated in a quiet language laboratory, in which it was possible to adjust the volume of the headphones individually. Subjects were told that they were going to take part in a perception experiment on Dutch intonation and were instructed to concentrate on the speech melody of the last syllable of each stimulus in the discrimination items. They were required to indicate whether they could hear any melodic differences between the stimuli and record their judgements in terms of the labels 'same' or 'different'.
Before the start of a session, the subjects heard 10 trial items in order to get accustomed to the task and to adjust the volume of the headphones to a comfortable listening level. Then the tape was stopped to enable subjects to ask questions. Subsequently, the subjects heard the items with gap size 40 msec. After a short break, the items with gap size 50 msec were presented for discrimination.

2. Results.

In this experiment, a total number of 5,735 observations were collected, i.e. 2,035 for the items in the ambiguous rise condition and 3,700 for the items in the rise condition 1. Most of the observations for rise condition 3 are missing due to technical problems with the language laboratory at the host institution. The judgements that were obtained for these items are regarded as unreliable because of frequent interruptions and are therefore not taken into account. The raw discrimination judgements for the stimuli are given in appendix A16 and A17.

2.1. The variables: Group, Gap, Rise Location and Alignment

In this section, the relationship between the independent variables rise LOCATION, GAP size, ALIGNMENT and average discrimination performance will be examined. Before going into the details of this analysis, the possibility of distinguishing different groups of subjects on the basis of similarities in their overall discrimination behaviour will be investigated.

2.1.1. Response Coincidence Analysis.

It should be recalled from the previous chapter that the main purpose of a response coincidence analysis is to establish whether it is appropriate to distinguish different groups of subjects on the basis of similarities in their overall discrimination behaviour. For this purpose, all the discrimination judgements of the subjects are compared to each other and this yields a response coincidence matrix, which is subsequently subjected to a hierarchical cluster analysis (Ward's method). For a more detailed description of this technique, the reader is referred to section 2.4.2. of chapter 4. The obtained response coincidence matrices for this experiment are given in appendix A18 and A19. The results of the cluster analyses are given in figures 5.5 and 5.6:

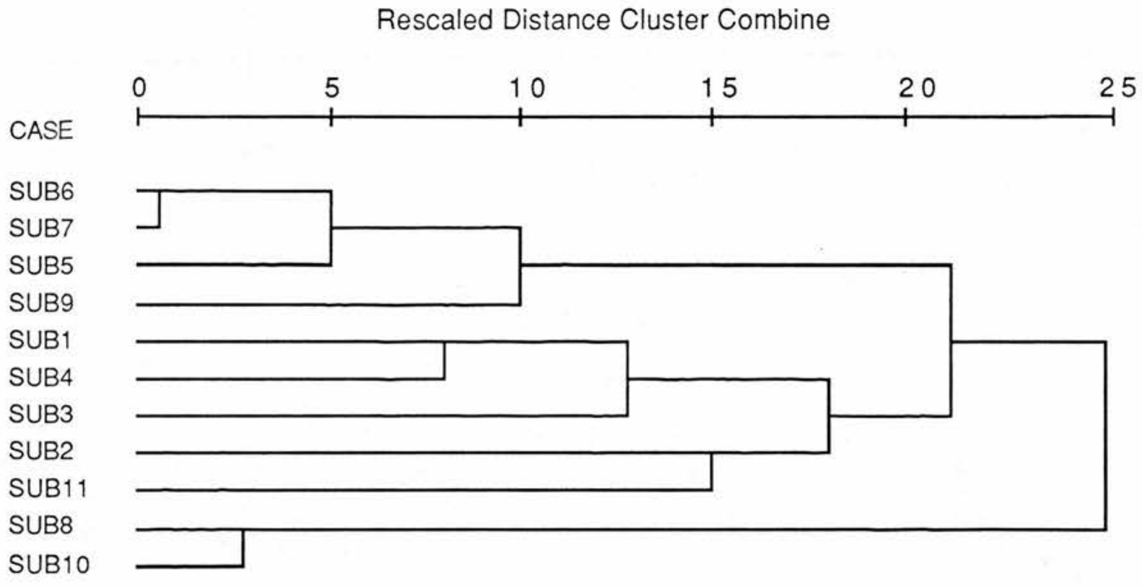


Figure 5.5: Dendrogram for informants in the ambiguous rise condition. Cluster 1 consists of subjects 8 and 10. Cluster 2 consists of subjects 6, 7, 5, 9, 1, 4, 3, 2 and 11.

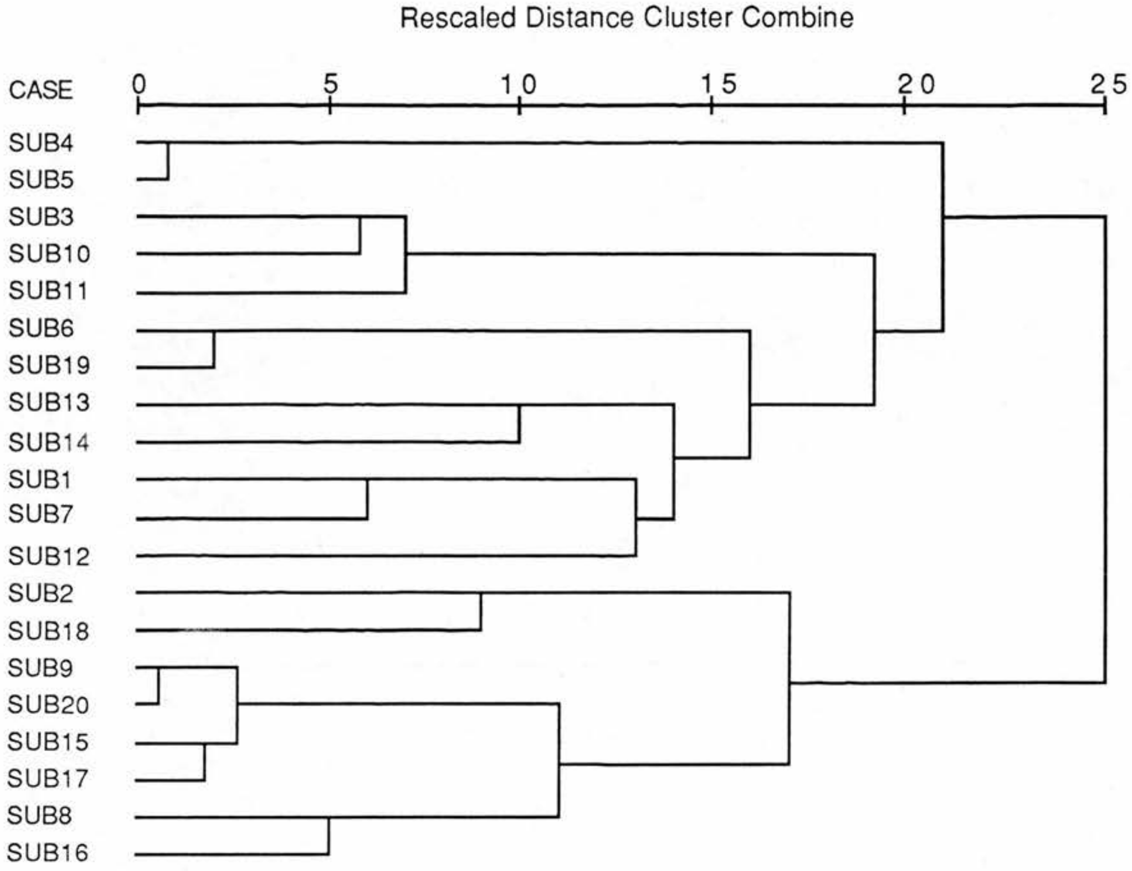


Figure 5.6: Dendrogram for informants in the rise condition 1. Subjects in cluster 1 are 2, 18, 9, 20, 15, 17, 8 and 16.

For consistency with the previous experiment, the highest clustering level is taken as the criterion to classify the subjects into separate groups. In order to define the discrimination criteria of the informants in each group, average item discrimination in the two subject clusters is calculated. The results are summarized in figure 5.7:

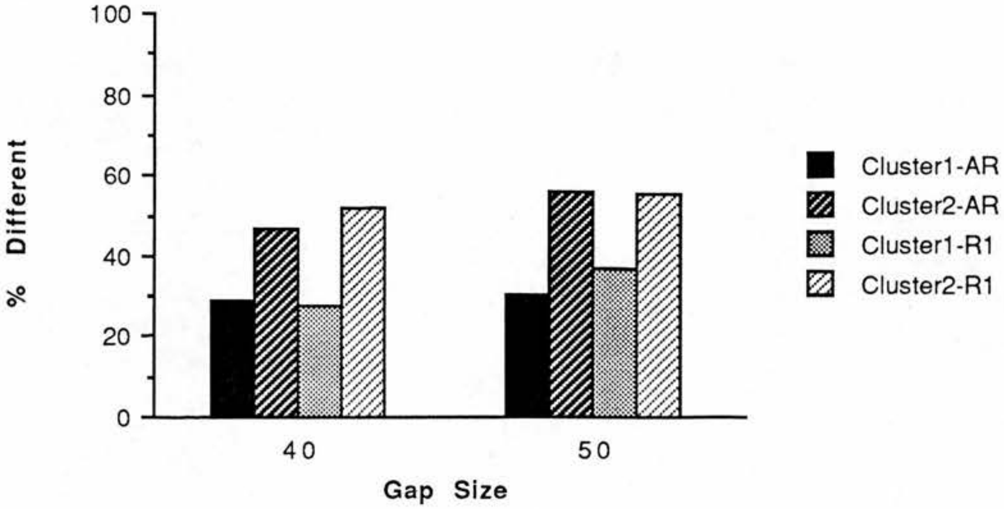


Figure 5.7: Average discrimination of the items in the ambiguous rise condition (AR) and rise condition 1 (R1) for the clusters identified by the cluster analysis. Responses are in terms of % different judgements.

It can be seen that the subject groups differ in the overall accuracy with which the items are discriminated. On average, the subjects in cluster 1 discriminate 31% of the items, whereas those in cluster 2 discriminate 52% of the items. Consequently, there is a justification to regroup the subjects as discriminators or non-discriminators on the basis of how well they discriminate the items of the experiment. By regrouping the subjects in this fashion, a new variable is introduced in the experimental design, i.e. subject GROUP.

Inspection of the total number of subjects in the two groups reveals that 71% of the informants belong to the discriminator group, whereas only 29% of the participants can be regarded as non-discriminators. These results are quite different from those obtained in the experiment on rise-alignment discrimination (Chapter 4), in which the largest group of subjects (60%) was found to consist of non-discriminators. In order to determine whether this difference in the number of subjects in the groups of this experiment and the previous one is statistically significant, a CHISQUARE was calculated. The independent variables in the calculation were subject GROUP and EXPERIMENT. The dependent variable was the number of subjects in the two groups of both experiments. A summary of these data is given in table 5.3:

	DISCRIMINATORS	NON-DISCRIM	TOTAL
Rise-alignment	38	58	96
Fall-alignment	21	10	31
Total	59	68	N=127

Table 5.3: Total number of subjects in the discriminator and non-discriminator groups of the experiment on rise-alignment differences and the present experiment.

Since there is only one degree of freedom in the table, the Yates Correction Factor was used (Hatch & Farhady, 1982). The CHISQUARE statistic indicates that the distribution of the number of subjects in the groups of the two experiments is significantly different ($\chi 2=6.38$, d.f.=1, $p < 0.025$).

A comparison of average item discrimination in the subject groups of the two experiments is given in table 5.4:

	DISCRIMINATORS	NON-DISCRIM	TOTAL
Rise-alignment	47	20	67
Fall-alignment	52	31	83
Total	99	51	N=150

Table 5.4: Average item discrimination (%) in the discriminator and non-discriminator groups of the experiment on rise-alignment differences and the present experiment.

These differences cannot be regarded as statistically significant ($\chi^2=0.63$, d.f.=1, n.s.). Nonetheless, it is interesting to observe that average discrimination in the present experiment is slightly better than in the experiment on rise-alignment differences. This observation holds for both the discriminator and non-discriminator groups.

The combination of these factors, i.e. that there are significantly more discriminators in this experiment (70% vs. 40%) and that the items are slightly better discriminated in both the discriminator and non-discriminator groups, can be taken as an indication that the items in this experiment are generally easier to discriminate in comparison to the items in the previous experiment.

2.1.2. Analysis of Variance.

As a second stage in the data analysis, an analysis of variance was carried out, which enables a comparison of average discrimination in the different experimental conditions. The independent variables in the analysis are 'rise LOCATION', 'subject GROUP', 'GAP size' and 'ALIGNMENT'. The dependent variable was taken to be the proportion of 'different' judgements for the discrimination items in each of the conditions. The result of the analysis is given in table 5.5:

ANALYSIS OF VARIANCE

	DIFF	Proportion of Different Judgements
by	ALIGNMENT	Location of the fall
	LOCATION	Rise Location in the Hat Pattern
	GROUP	Subject Group
	GAP	Gap Size

Source of Variation	Sum of squares	DF	Mean Square	F	Sig.
Main Effects	19.323	20	.966	14.900	.000
GROUP	12.239	1	12.239	188.744	.000
GAP	.499	1	.499	7.696	.006
LOCATION	.382	1	.382	5.897	.015
ALIGNMENT	6.557	17	.386	5.948	.000

None of the 2-, 3- or 4-way interactions is significant.

Source of Variation	Sum of squares	DF	Mean Square	F	Sig.
Explained	27.210	143	.190	2.934	.000
Residual	63.028	972	.065		
Total	90.238	1115	.081		

1116 cases were processed

Table 5.5: Results of the analysis of variance by means of the SPSS-X software. The independent variables are LOCATION, GROUP, GAP and ALIGNMENT. The dependent variable is the proportion of different judgements.

The analysis of variance indicates that all the independent variables in the experimental design are statistically significant main effects. It also emerges that there are no significant higher order interactions, so that the discussion of the results can concentrate on the main effects only. The relationship between GROUP and discrimination is summarized in figure 5.8:

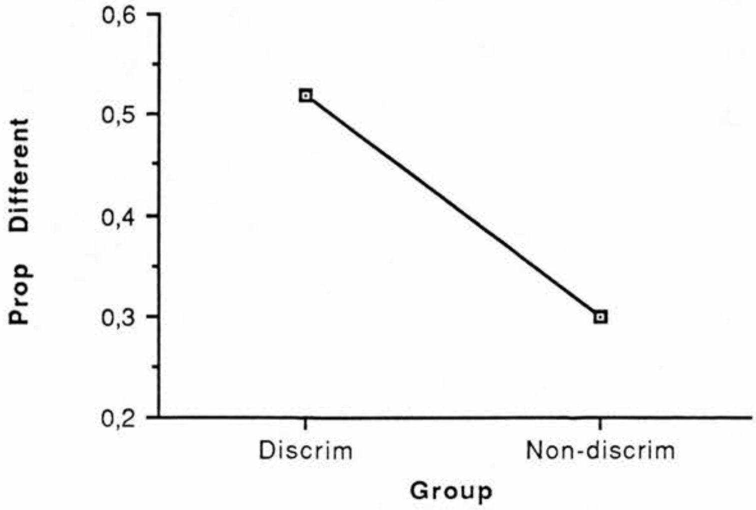


Figure 5.8: Average proportions of 'different' judgements as a function of subject group.

This constitutes further statistical justification for distinguishing these groups of subjects, in that the discriminator group generaly discriminates the items better than the non-discriminator group. This difference is statistically significant ($p < 0.001$). The relationship between GAP and discrimination is given in figure 5.9:

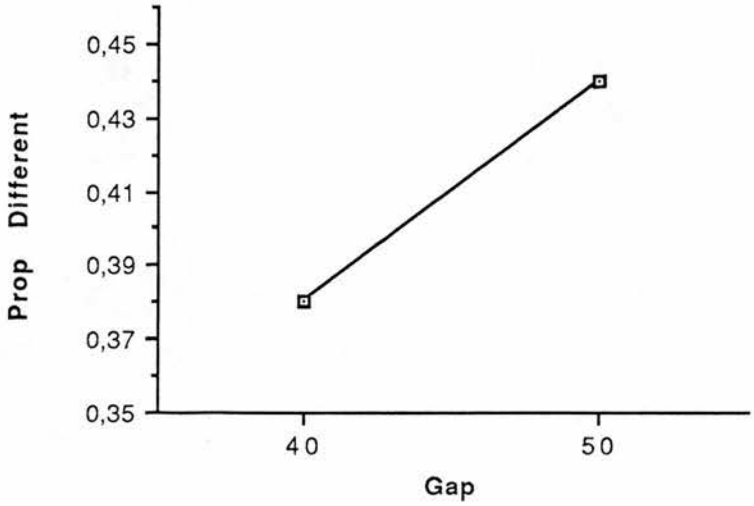


Figure 5.9: Average proportion of 'different' judgements as a function of gap size.

As far as this variable is concerned, a statistically significant (p = 0.010) relationship is observed between GAP size and average discrimination, i.e. the items with gap size 50 msec are generally better discriminated than those with a 40 msec gap between the endpoints of the falls in the stimuli. Although a similar relationship was found in the experiment on rise-alignment differences, the effects in both experiments are not entirely identical. In the previous experiment, the relationship between GAP and discrimination was only significant in the BA (late-early) presentation order, whereas in the present experiment the effect clearly manifests itself in the AB (early-late) order of presentation.

Next, it can be observed that the LOCATION of the rise in the hat pattern stimuli has a statistically significant effect on average discrimination. This suggests that the items are better discriminated when the contours contain a rise of type 1. The effect is however not very large, i.e. 45 % vs. 44 %.

Finally, the main variable of this experiment, i.e. ALIGNMENT, turns out to be statistically significant. The average discrimination function, which was obtained by averaging all the judgements for the discrimination items with the same fall location in the first stimulus is given in figure 5.10:

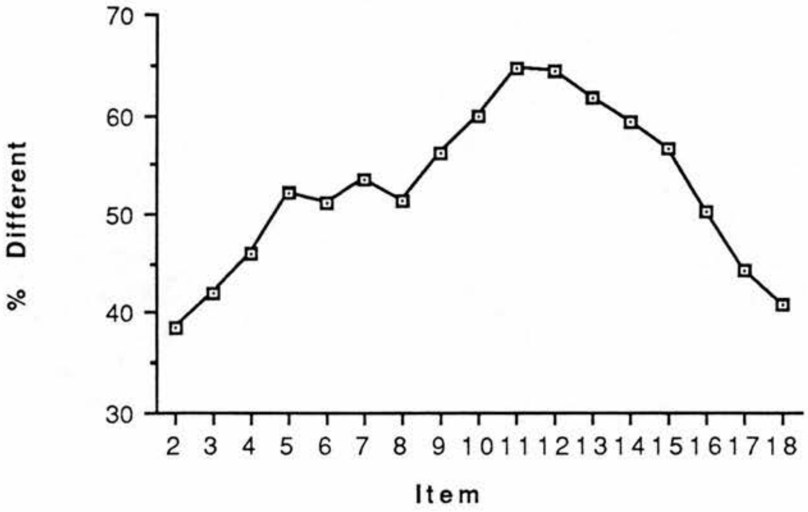


Figure 5.10: % 'different' judgements as a function of fall alignment. The curve is smoothed(1) by two points.

This average discrimination curve is characterized by a discrimination peak at item number 11, while the items at the extremes of the continuum are not discriminated. The fact that none of the higher order interactions is significant, indicates that the distribution of 'different' judgements as a function of ALIGNMENT is essentially identical in all experimental conditions.

3. Discussion.

The results of this experiment on the discrimination of fall alignment differences indicate that the four experimental variables in the design are statistically significant. As far as GAP size is concerned, a significant increase in average discrimination is observed with increasing physical differences along this dimension. Although this could reasonably be expected on a purely psychophysical basis, it is worth pointing out that this effect is not significant in the early-late presentation order of the previous experiment, in which a correlation with increasing gap size is only observed in the late-early presentation order.

---

(1) In the curve-smoothing procedure, the discrimination judgements in adjoining items of the continuum were added and divided by two. The statistical procedures are NOT based on the smoothed judgements, but on the absolute total of 'different' judgements on each item.

Regarding rise LOCATION in the stimuli of the items it is found that there is a small but significant relationship between this parameter and average item discrimination. Finally, a response coincidence analysis suggests a classification of participants into two groups as discriminators and non-discriminators. Unlike the previous experiment, in which the largest number of subjects belong to the non-discriminator group, most of the informants in this experiment (70%) are classified into the discriminator group. Moreover, it is observed that average item discrimination in this experiment is slightly better than in the experiment on rise-alignment differences, both for discriminators and non-discriminators.

The main purpose of this experiment is to establish the perceptual relevance of fall-alignment differences to the characterization of the hat pattern in Dutch. Several aspects of the data have to be discussed in this perspective. First and foremost, it is observed that there is a significant discrimination peak in the discrimination function. If we take the average discrimination function (figure 5.10) as a starting point, it can be seen that this peak is located at item number 11, where discrimination reaches a maximum of 69%. A further characteristic of this function is that the informants do not discriminate the items representing the extremes of the fall-alignment continuum. A discrimination function of this kind is highly consistent with functions obtained in experiments on the categorical perception of segmental features of speech, such as voice onset time. It essentially indicates that equal physical differences in fall-alignment are not discriminated uniformly throughout the continuum, but rather that discrimination is discontinuous. This can be accounted for in terms of the phoneme-boundary effect (Wood, 1976), which indicates that discrimination is significantly better when subjects are asked to discriminate stimuli which represent different categories. When subjects are required to discriminate stimuli representing the same category, they fail to notice any differences. This suggests that the alignment continuum is perceived in terms of two categories of falls depending on their location in the experimental syllable.

As to the precise location of the category boundary in the alignment continuum, the discrimination peak provides essential information. In this continuum, the peak is located at item number 11, the first stimulus of which contains a fall that is located entirely in the syllable's vowel nucleus. The endpoint of the fall in the second stimulus is located in the postvocalic consonant. This seems to indicate that the category boundary in the continuum coincides with the segment boundary between vowel nucleus and postvocalic consonant. When the endpoints of the falls in the stimuli are located on either side of this segmental boundary, discrimination remains relatively high. This is the case for items 11, 12 13 and 14 and this explains the resistance of the curve to drop down immediately after the discrimination peak has been reached. The proposed location of the category boundary is consistent with the classification of falls into a late (A) and a very late (C) category in the GDI. The fact that the extremely late fall (B) does not materialize in the discrimination function obtained here is not contrary to expectation, since falls of this kind are assumed to occur between syllables. It is clear that such a context is not provided in the stimuli of the experiment.

It should be emphasized that the results of this experiment do not allow us to conclude that the two categories of falls are distinguished on the basis of fall alignment alone, since there is the additional perceptual cue of fall 'completeness'. In the category representing falls of type A, the falls are complete in that F0 falls to the lower declination line in the syllable nucleus. In the second category representing falls of type C, the movements are incomplete in that they are partially located in the voiceless postvocalic consonant.

The second finding of this experiment relates to the hierarchical relationship between the rise and the fall in the hat pattern. In the GDI, it is postulated that there is a relationship between the location of both movements in the pattern: if the rise occurs late, the fall occurs late as well. If the rise is aligned early, the fall is located early. This conclusion certainly seems justified on the basis of a systematic investigation of a large set of pitch contour stylizations in Collier (1972). It is clear from the perceptual data presented here that the location of the rise in the pattern has a small but significant effect on fall discrimination. In this perspective, it is unfortunate that the results for the discrimination items with a very late rise are missing, since these could have provided a more complete picture of this relationship.

The final observation in this experiment relates to the subject groups and their average discrimination of the items. It is found that there are substantially more discriminators than non-discriminators in this experiment and that this is the opposite of what was found in the previous experiment. In addition, all the items of this experiment are generally better discriminated than those of the previous experiment. This seems to suggest that the items of the fall-alignment continuum are easier to discriminate than those of the rise-alignment continuum, especially when it is considered that the 80% of the items in the rise-alignment task contained physical differences equal to or greater than those in the fall-alignment experiment. All these factors taken together suggests that informants are more sensitive to variations in falling pitch movements than to variations in rises. This can be accounted for in terms of categories of pitch movement alignment, in that the single-rise category spans a larger physical area of the continuum than the two fall categories. Consequently, physical differences in rise-alignment may be larger since there is less danger of spanning a category boundary. In falls, however, the same segmental stretch is divided into two categories, so that the sensitivity of informants is increased.

4. Conclusions.

This experiment provides substantial perceptual evidence that informants' discrimination of equal physical differences in fall-alignment is not uniform, but reflects a categorization of falls into two major classes. The first category is constituted by falls which are entirely situated within the syllable nucleus (type A), whereas the endpoints of the falls in the second category are situated in the postvocalic consonant (type C). This is entirely in agreement with the classification of falls in the GDI. As a result, a specification of fall-alignment is essential to the characterization of the hat pattern, in that it has to be located entirely in the syllable nucleus. If not, a perceptually different pattern is involved, i.e. the cap pattern, in which the fall is aligned as late as possible and consequently is incomplete.

It is also shown that a hierarchical relationship between rise location and fall-alignment is substantiated. Finally, greater sensitivity of informants to fall-alignment differences is found to be consistent with the categorization of falls into two categories.

CHAPTER 6

The Discrimination Threshold of Pitch Movement Alignment.

0. Introduction.

In the previous experiments, the categorization of pitch movement alignment in the hat pattern was investigated for both rises and falls. The results of the discrimination experiments discussed in chapters 4 and 5 suggest that informants are more sensitive to fall-alignment differences than to differences in rise-alignment. It was hypothesized that this differential sensitivity may have arisen from the number of tonal categories that underlie the alignment continua. The rises that are associated with a syllable constitute a single alignment category and the exact location of the rise is not perceptually relevant, provided it stays within the syllable boundaries. Consequently, it can be assumed that the perceptual system is fairly robust regarding alignment differences in rises. As far as falling pitch movements are concerned, the perceptual system has to be more finely-tuned: since the syllable duration represents two tonal categories, the distance needed to cross a category boundary in necessarily smaller. In order to investigate the hypothesis that informants are more sensitive to fall-alignment than to rise-alignment in the hat pattern, an experiment was carried out to establish the just noticeable difference of pitch movement alignment in this pattern. If informants are indeed more sensitive to fall-alignment than to rise-alignment, it is to be expected that the JND of fall-alignment is considerably smaller than that of rise-alignment.

1. JND of Pitch Movement Alignment in the Hat Pattern.

1.1. Experimental Design.

In order to establish the JND of pitch movement alignment in the hat pattern, two fixed-standard 2IAX discrimination experiments were carried out, in which informants were asked whether they could discriminate between pairs of stimuli that were taken from a pitch movement alignment continuum. In the first part of the experiment, the informants judged stimuli from a rise continuum; in the second, stimuli from a fall continuum were used.

1.1.1. Stimulus Preparation.

The stimuli for this experiment were prepared by means of the same technique of modified resynthesis described in the methodological section of chapter 4. The utterances that were chosen as the basis for the stimuli were the same as in the experiments on the discrimination of rise- and fall-alignment, i.e.:

(1) Renaat is ziek (Renaat is ill)

(2) Renaat is in Parijs (Renaat is in Paris)

The stimuli derived from (1) were used in the rise-alignment task, while those derived from (2) served in the fall-alignment discrimination task. Although the use of stimuli that are segmentally different has clear disadvantages, we aimed to keep the experimental conditions similar to those of the experiments in chapters 4 and 5. Both sentences were read by a native speaker of Dutch, who realized each utterance with a flat hat pattern. One delivery of each sentence was chosen which sounded natural and fluent in all respects.

These utterances were digitized, analysed acoustically and provided with a standard flat hat pattern, the specification of which conformed to values given in experiment 1. The pitch movements in the experimental syllable were aligned in such a way that their onsets coincided with the onset of the prevocalic consonant. For the rise continuum, the rise coincided with the onset of the syllable 'NAAT' in 'reNAAT'. In the fall continuum, the fall started at the onset of 'RIJS' in 'paRIJS'. Both duration and excursion size of the movements in the experimental syllables were identical, that is, 100 msec and 5 semi-tones.

The actual stimuli were then derived from these reference contours. In stimuli for rise alignment, the rise in the stressed syllable of 'reNAAT' was shifted to the right in steps of 10 msec. Hence a total of 22 stimuli were obtained. In the stimuli with falling movements, the falls were shifted to the right through the syllable 'paRIJS', which resulted in 19 stimuli. The exact location of the pitch movements in the stimuli is given in appendix A20, and is illustrated here schematically in figure 6.1:

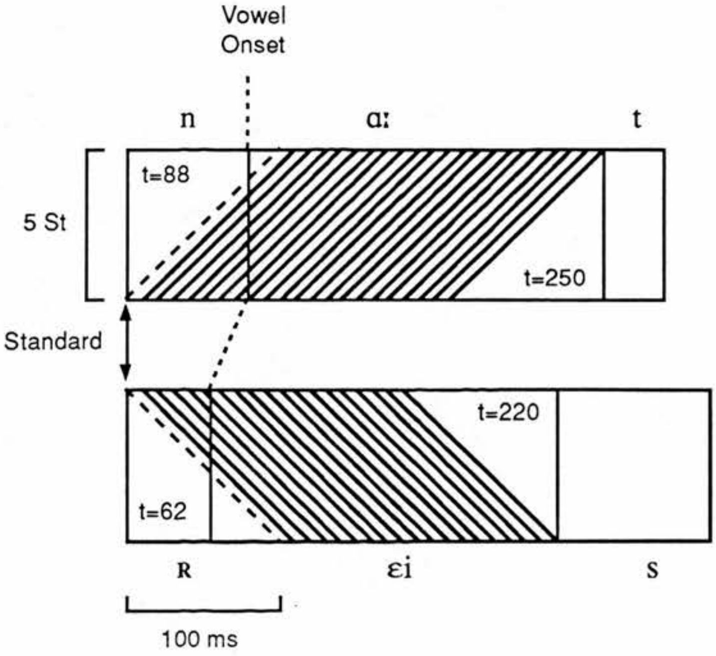


Figure 6.1 : Schematization of the range of variation in the stimuli for rise alignment (top) and fall alignment (bottom). Vertical lines indicate the segment boundaries and durations are specified in milliseconds. The dashed lines represent the alignment of the standard in the experimental syllable of the reference stimulus.

1.1.2. Item Construction.

The stimuli from the rise continuum and those of the fall continuum were combined into AX discrimination items in such a way that the reference stimulus was paired with every other stimulus from the continuum. As a result, gap sizes for rise-alignment were obtained ranging between 10 and 210 msec, whereas those for fall-alignment ranged between 10 and 180 msec. In all the items, the reference stimulus was presented first and each item occurred five times in the test. This yielded 105 items (21 x 5) for rise alignment and 90 items (18 x 5) for fall alignment.

Although it is common practice in experiments of this nature to include a series of AA or BB items, i.e. items in which the stimuli have identical pitch movement locations, it was decided to omit these from the test. This was motivated by the need to reduce informants' listening time. This precludes an estimate of subjects' response bias as well as an estimate of chance discrimination.

1.1.3. Test Tapes.

A separate test tape was recorded for the rise and fall discrimination task, the items being randomized 10 times. The inter-stimulus interval was 500 msec, whereas the inter-item interval was 3000 msec. No identifiers were provided between the items, but after each set of 10 items, the informants heard an orientation signal.

1.1.4. Informants.

A total of 14 informants took part in the experiment. They were Belgian and Dutch visiting students at the University of Aarhus (Denmark). They were all native speakers of Dutch and participated on a voluntary basis.

1.1.5. Procedure.

The subjects were seated in a quiet language laboratory, in which the volume level of the headphones was individually adjustable. They were told that they were going to take part in an experiment on Dutch intonation, which was aimed at establishing which intonational differences are perceivable by native speakers of Dutch.

One of the problems that have to be recognized relates to the instructions given to the informants in performing the task. Since it is the aim to establish the JND of pitch movement alignment, it seems logical to have informants judge the location of the movement in the second stimulus with respect to the first in terms of the labels *later* and *earlier*. Informal experimentation has shown, however, that informants have considerable difficulty in handling these concepts reliably. Therefore, it was decided not to specify the acoustic dimension explicitly in the experimental task. In the first session, the informants were instructed to concentrate on the melodic characteristics of the word *Renaat* in each stimulus pair and were asked to indicate whether they could hear a melodic difference or not. Their judgements were given on a scoring sheet in terms of *same* or *different*. The task in the second session was identical, except for the fact that informants were instructed to listen for melodic differences on *Parijs* in the stimuli of each item.

Before the start of the experiment, the subjects were presented with 10 trial items to get used to the task and adjust the volume to a comfortable listening level. Between both sessions, there was a 10 minute break. Total duration of the experiment was approximately 60 minutes.

1.2. Results.

In this experiment, a total of 2730 discrimination judgements were obtained, i.e. 1470 for rise discrimination and 1260 for fall discrimination. The raw discrimination judgements are given in appendix A 21. The discrimination curves are given in figure 6.2:

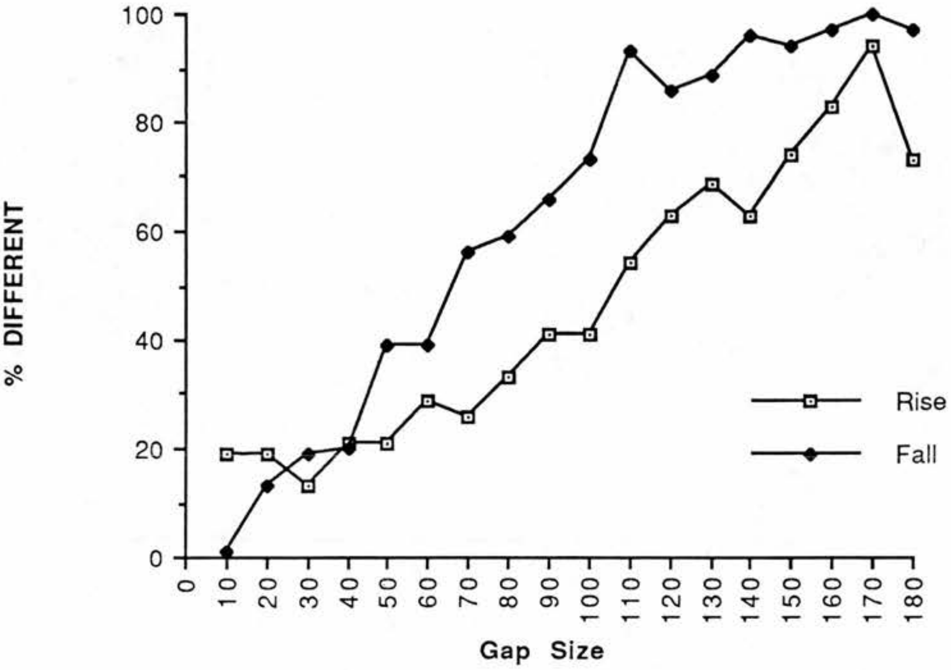


Figure 6.2: % 'different' judgements as a function of gap size (msec) for the items in the rise and fall condition. Each datapoint represents 70 observations.

Visual inspection of the discrimination results in figure 6.2 indicates that the percentage of 'different' judgements increases as a function of gap size. At the same time a striking difference between the discrimination judgements in the rise and fall condition is observed, in the sense that the discrimination curve is markedly steeper in the fall condition than in the rise condition: as gap size increases, the percentage of 'different' judgements increases more rapidly in the fall condition than in the rise condition.

As a first step in the analysis of the results, the just noticeable difference of pitch movement alignment was measured in the two experimental conditions for the individual subjects. The just noticeable difference is defined here as the stimulus difference which is noticed 50% of the time (Guilford, 1954). These measures were obtained by means of the method of average z-scores (Woodworth & Schlosberg, 1954). In this method, the observed proportions of 'different' judgements for the gap sizes are converted into z-scores. Subsequently, the mean is taken for the lower half of the gap sizes and the mean of the corresponding z-scores is assigned to it. The same is done for the upper half of the gap sizes. These derived z-scores are marked on a graph in which z-scores are plotted as a function of gap size and a straight line is drawn through them. The result of this procedure is illustrated in figure 6.3 by means of the fall-discrimination data from a single speaker:

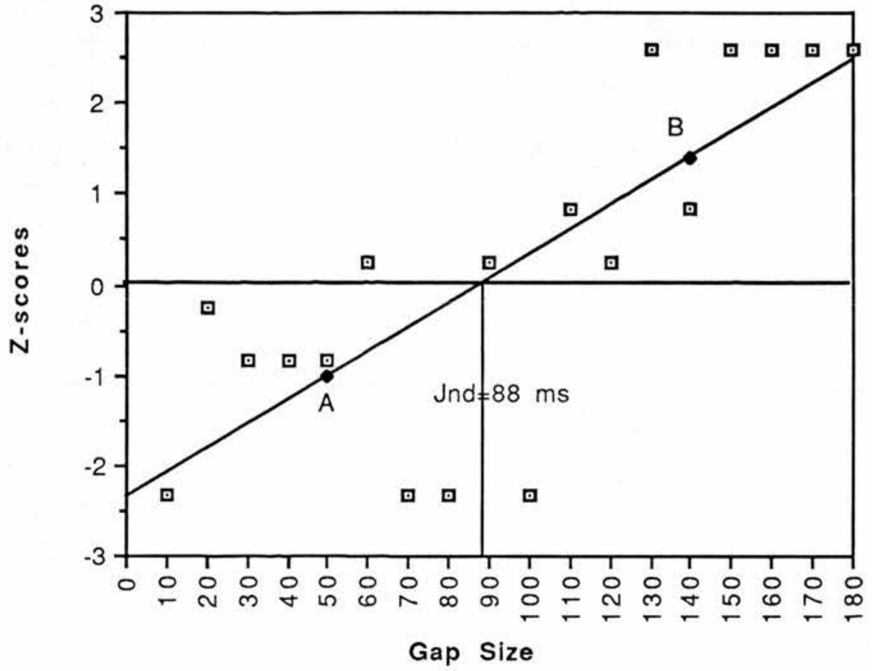


Figure 6.3 : Scatterplot of z-scores as a function of gap size in the fall-alignment continuum with 18 gap sizes. 'A' represents the mean z-score for mean of the lowest 9 gap sizes, while 'B' repesents the mean z-score for the highest 9 gap sizes. The JND corresponds to the gap size where the straight line through A and B crosses the horizontal line where z=0.

The point where this line crosses a horizontal line where z=0 (i.e. p=0.50) is taken as the measure of the JND. This method was chosen, since it relies on all the discrimination data obtained in the experiment to estimate the JND. The method is defensible because the relationship between the z-scores and gap sizes is linear, so that a straight line can be fitted through the data, whereas the relationship between the observed proportions and gap sizes is of a sigmoidal nature.

The JND's obtained by this procedure were subsequently submitted to a oneway analysis of variance. The independent variable in the analysis was the TYPE of pitch movement (rise vs. fall). This analysis gives an F-value of 13.254 (d.f. = 1), which is significant at p < 0.001. The mean JND for rise-alignment is 100.77 ms, whereas this for fall-alignment amounts to 68.07 ms. The observed difference in JND between the rise and fall condition is consequently 32.7 ms.

1.3. Discussion.

As could have been expected, the data presented above show a straightforward relationship between gap size and discrimination performance: the larger the gap size, the better discrimination. Although this is an almost trivial statement of the facts, there is one aspect of the data which deserves special attention, i.e. the significant difference in overall discrimination scores in the rise and fall condition. The percentage of different judgements increases more rapidly in the fall alignment (FA) than in the rise alignment (RA) condition, resulting in significantly different JND's for rise and fall alignment discrimination. These figures seem to provide support the hypothesis that informants are more sensitive to fall alignment differences than to differences in rise alignment.

As for the JND for rises, Collier (1983) suggested a value of 50-70 msec depending on the spectral characteristics of the experimental syllable. The present data however indicate that the JND is in fact 101 msec, which is considerably larger than expected. This value assumes that it is legitimate to make abstraction from the order in which the stimuli are presented in the discrimination items and the location of the standard rise in the experimental syllable, since all the stimuli were presented in AB (early-late) order and the standard was fixed at the onset of the accented syllable. The JND for falls is found to be 68 msec, which corresponds well with the value suggested in Collier (1983).

Although the differential threshold seems to indicate that informants are more sensitive to differences in FA than in RA, it cannot conclusively be inferred from the experiment that this differential sensitivity relates to the occurrence of the movements in the hat pattern. Due to the limitations of the experimental design, a critical appraisal of various factors that may have caused the differential threshold has to be undertaken.

1.3.1. Stimulus Design Factors.

It was mentioned in section 1.1.1. that the pitch movements of the reference stimulus are aligned to the onset of the experimental syllable, i.e. the onset of the movement coincides with the onset of the prevocalic consonant. Both the experimental rises and falls have an equal duration of 100 msec. The durations of the prevocalic consonants in the experimental syllables are however different, with the result that the endpoint of the reference rise is located earlier in the vowel nucleus (Vowel Onset +12 msec) than that of the reference fall (VO + 38 msec). In addition, the durations of the experimental vowel nuclei differ as well: 250 msec in RA versus 220 msec in FA. The interaction between these different parameters is schematized in figure 6.4:

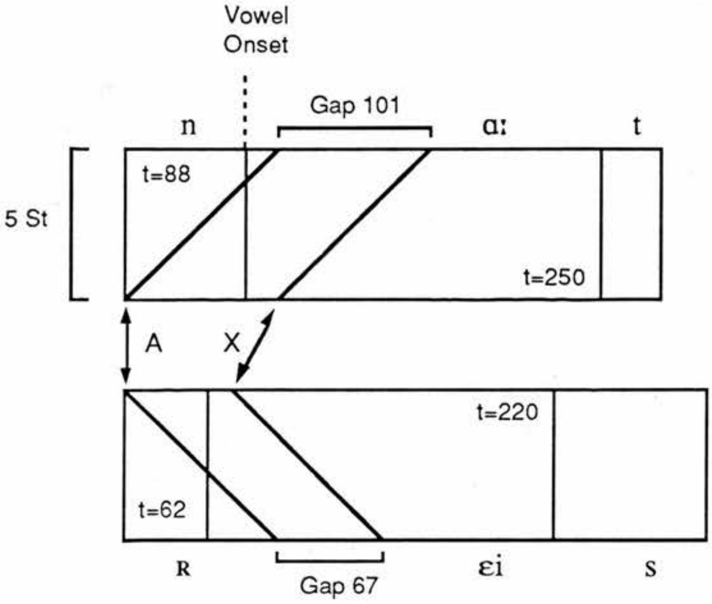


Figure 6.4: Durational characteristics of the experimental syllables and alignment of the pitch movements in the stimuli of the items at the discrimination threshold for the rise (top) and fall conditions (bottom).

In figure 6.4, it can be noted that the endpoints of the movements in the second stimulus (X) of the threshold items are situated at proportionally the same location in the vowel nucleus of the experimental syllable in both the RA and FA conditions, i.e. at 48% of the total vowel duration. If this location is in some way highly critical to discrimination, it follows that a bigger absolute gap size is needed in the rise condition to reach this critical boundary than in the fall condition. This results logically from the combined effect of the unequal duration of the vowel nuclei in both conditions and the different locations of the endpoints of the movements in the standard stimulus with respect to vowel onset.

Evidence for the importance of the middle portion of the vowel nucleus was provided in Rossi (1971, 1978) who showed that this area is critical to the perception of height of pitch transitions, in the sense that informants match perceived pitch of a transition to a pitch level corresponding to the middle third of the vowel stimulus. Informants react in the same fashion to rises and falls. Similar results were reported in Nabelek et al. (1970).

This *boundary hypothesis* has quite important consequences for the relationship between gap size and discrimination, since it suggests that it is not the gap size in absolute terms which is the decisive factor in the discrimination of alignment differences, but rather whether the movement endpoint in the second stimulus is located beyond a critical boundary or not. This critical boundary is assumed to be situated in the middle of the syllable nucleus. The hypothesis thus predicts that much smaller gap sizes than the thresholds suggested above can be perceived, provided that the pitch movements in the two stimuli are located on different sides of this boundary.

It should be pointed out that the boundary proposed here has to be of a psycho-acoustic nature, since a linguistic boundary is not consistent with the discrimination results reported in chapter 4 and 5. The discrimination experiment for fall-alignment provided a clear indication of the existence of such a category boundary: it is however located at the very end of the syllable nucleus and not in the middle of the vowel. The discrimination data for rise-alignment in chapter 4 provide no evidence of a category boundary at all. In the light of these data, the boundary hypothesis has to assume a psycho-acoustic boundary rather than a linguistic one.

The second factor that has to be taken into consideration here, relates to the spectral characteristics of the experimental syllables in the rise and fall conditions. It is suggested in Collier (1983) that there is a relationship between discrimination of pitch movement alignment and the spectral characteristics of the experimental syllable, in that alignment discrimination is better in a spectrally unstable area than in a spectrally stable stretch of speech. The results obtained in this experiment can be interpreted in this light, since the spectral characteristics of the experimental syllables in the rise and fall condition are quite different. In the rise condition, the alignment continuum was implemented on a spectrally stable [a] monophthong, whereas in the fall condition, an [ei] diphthong is used which is characterized by continuous spectral change. Consequently, the better discrimination of fall alignment can be an artifact of the complex relationship between spectrum and pitch discrimination. This will be called the *spectral hypothesis.*

Although Collier's observations are consistent with the observed differential threshold in this experiment, it should be pointed out that recent research indicates that the relationship between pitch discrimination and the spectral characteristics of the segmental string is different from that proposed by Collier. In House (1990), the spectral constraints hypothesis is put forward, which holds that the discrimination of pitch features is better in spectrally stable stretches of speech:

> As the complexity of the signal increases, pitch sensitivity decreases. We can then speculate that when the perceptual mechanism is maximally loaded with the task of resolving spectral information and rapid changes in intensity (...) its capacity of resolving fundamental frequency movement is decreased. Sensitivity to tonal movement would therefore be greatest during areas of relative spectral stability and least during areas of spectral and intensity change (p. 34).

The results of several pitch discrimination experiments reported in House (1990) provide substantial support for this hypothesis. If the spectral characteristics of the segments are to account for the differential threshold, the spectral constraints hypothesis predicts that informants should be least sensitive to fall-alignment, since the falls are implemented on a diphthong. This is clearly not the case, so that in the light of House's data, the spectral hypothesis is not supported.

The third stimulus design factor relates to the presentation order of the stimuli in the discrimination items. In this context, it can be assumed that physical changes in pitch movement alignment correspond perceptually to changes in the overall pitch level of the experimental syllable. The rationale behind this assumption is based on results in Nabelek et al. (1970). They carried out a matching experiment in which informants were asked to match the perceived pitch of a frequency transition to a set of reference stimuli with a steady pitch level at various frequencies. The frequency transition did not take place over the whole duration of the stimulus, so that the rising transitions were preceded by a portion of steady state low frequency and followed by a steady state high F0. In the falls, the reverse applied. Moreover, the transitions were shifted through the stimulus, which in effect created an alignment continuum. Nabelek et al concluded that:

> When the frequency transition is small (...) the pitch is evaluated according to the position of the transition. When the transition occurs with some delay, the judgements shift toward the initial frequency, i.e. toward the lower frequencies for rising changes or toward the higher frequencies for falling changes (p. 540)

This indicates that delaying rises in a continuum corresponds perceptually to a lowering of the pitch, whereas postponing falls causes the pitch of the syllable to rise.

In this perspective, the discrimination items in the rise and fall conditions are symmetrical in terms of pitch movement alignment in that the pitch movement of the second stimulus always occurs later than the pitch movement in the first stimulus of each item. The items are however clearly asymmetrical in terms of perceived pitch height associated with the movements in the stimuli. In the rise condition, the early-late presentation order corresponds to a lowering of the overall pitch level of the experimental syllable, whereas in the fall condition, the perceptual effect of the early-late presentation order is reversed: the overall pitch level is raised.

Interpreted in this way, the threshold differences found between the rise and fall condition suggest that informants are more sensitive to pitch raising than to pitch lowering. This is entirely consistent with my own results on the discrimination of pitch excursion size in the pointed hat pattern in Dutch, which indicate that raising an F0 peak is discriminated significantly better than lowering the peak by an equal amount. This is also consistent with 't Hart (1981). He found that the JND for pitch movement excursion size is smaller for rises than for falls, which may indicate that informants are more sensitive to pitch raising than to pitch lowering.

It can be argued against this interpretation, that the differential sensitivity to pitch height variations just reflects a compensation strategy due to the declination phenomenon: the informants naturally expect the pitch level of the experimental syllable in the second stimulus to be lower than that of the first stimulus, with the result that lowering the peak in the second stimulus is not very noticeable. This is the case in the items of the rise condition. Raising this level on the other hand is in conflict with what would be expected by declination, and can be argued to be more easily perceivable. This explanation can only work if it is assumed that the declination phenomenon operates across stimulus boundaries, which is contradicted by results in Leroy (1984). She conducted an experiment in which subjects had to judge the height of two pitch peaks in different utterances. Her results suggest that:

> (...) listeners do not compensate for declination when judging the relative height of successive pitch peaks that belong to separate utterances (p. 61)

The final stimulus-design feature that has to be considered is the overall placement in time of each alignment continuum in the stimuli. In the stimuli for rise-alignment, the continuum is located at the beginning of the utterance, while in the stimuli for fall-alignment, the continuum is located at the end of the utterance. As a result, the independent variable of pitch movement TYPE (rise vs. fall) is confounded with the differential placement in time of the two alignment continua. In this perspective, the differential threshold may indicate that informants are more sensitive to alignment differences at the end of utterances (falls) than at the beginning of utterances (rises). It should be emphasized that this difference in placement of the continuum in the stimuli cannot be avoided if the flat hat pattern is taken as the research configuration, since it is its fundamental characteristic that the rise occurs earlier in time than the rise. Consequently, an experiment in which the placement of the rise and fall continua in the stimuli is held constant is practically impossible to design using the hat pattern as a configuration, without introducing all kinds of other variables of which the influence on pitch discrimination is impossible to predict.

To my knowledge, the potential effects of the placement of the alignment continuum on discrimination has never been investigated systematically, but it was used as a controll variable in an alignment categorization experiment reported in House (1990). In this experiment, the placement in time of a fall alignment continuum was manipulated as a controll variable. The results of this experiment indicate that the identification curves are identical in both placement conditions. This suggests that this variable is unlikely to cause significant differences in discrimination.

1.3.2. Linguistic Factors.

The hypotheses that are formulated in the previous section suggest that the differential threshold in the discrimination of pitch movement alignment originates from psycho-acoustic factors. An alternative explanation can be attempted if it is explicitly assumed that the differential threshold relates to underlying linguistic differences.

The first possibility to be considered here, relates to the phonological status of the accents which are realized by both types of pitch movements. In the case of rises, the accent realized by the movements has a prenuclear status, whereas the falls correlate with a nuclear accent. Due to the phonological dominance of nuclear over prenuclear accents, it is reasonable to assume that informants are more sensitive to variations in phonetic cues in a nuclear accent position. To my knowledge, there is no systematic perceptual evidence that this may be the case, but it presents itself as an interesting possibility. An interpretation of the data in this light is clearly problematic for the GDI, since it postulates an equal status of the movements associated with accents in a linguistic perspective: they are all prominence-lending with no further hierarchical subclassifications. Models which do postulate a hierarchical structure of prominence are clearly better equipped to relate differential sensitivity to underlying phonological structure.

The second possibility, which to some extent is related to the first, is that the JND may be related to the size of the tonal category associated with the alignment category. The discrimination data in chapter 4 suggest that there is only one category of rise-alignment in the Dutch hat pattern. All the rises, the endpoints of which are located in the syllable nucleus or the postvocalic consonant of the experimental syllable, are regarded as instances of the same alignment category. If the duration of the syllable nucleus and the postvocalic consonant is taken as an indication of the physical size of the rise-alignment category, the size of the rise-category is 308 msec.

The results in chapter 5 indicate that falls aligned with the syllable nucleus and the postvocalic consonant can be classified into two categories. The endpoints of the falls in the first category, which is involved in this experiment, have to be located anywhere before the end of the syllable nucleus. Here, the duration of the syllable nucleus only provides an indication of the physical size of the category, which is 220 msec. The calculation of the JND as a proportion of the category size reveals that rises are just noticeably different if the gap size amounts to 33% of category size. For falls, this proportion is 31%. From this, it can be concluded that the difference in JND of both types of movements is negligible if it is expressed as a proportion of alignment-category size and this account will be referred to as the *category-size hypothesis*, which explicitly postulates that the JND is a proportion of the tonal category size. If the JND's are expressed as proportions of the physical size of the alignment categories that were established in the previous chapters, the JND's are almost identical. This may well reflect an influence of implicit linguistic knowledge about the size of the tonal categories on discrimination.

The hypothesis that the JND is a proportion of the tonal category size is consistent with the results of the previous experiments and suggest that the sensitivity of the human perceptual mechanism is determined by implicit linguistic knowledge. The perceptual mechanism is finely-tuned as far as pitch differences which signal linguistic differences are concerned, as in the case of fall-alignment. The perceptual mechanism is more robust regarding pitch differences which do not signal linguistically meaningful distinctions, such as rise-alignment in the hat pattern. This fits in well with 't Hart (1979a), who observed that the perceptual tolerances in the perception of intonation are not uniform:

> (...) the perceptual tolerances which make stylisation possible at all, do not seem to be uniform, that is, to our experience, one has to be very precise on some places, whereas on other places it does not matter very much ('t Hart, 1979a: 377)

The category-size hypothesis proposed here accounts for this non-uniformity of pitch discrimination by relating perception to linguistic structure.

1.4. Conclusions.

The discrimination data presented in this experiment indicate that informants are more sensitive to fall-alignment than to rise-alignment, which is reflected in the discrimination thresholds for pitch movement alignment. The JND for fall-alignment is 68 msec, and for rise-alignment it is 101 msec. Due to the limitations of the experimental design, this differential sensitivity cannot be unambiguously related to the type of pitch movement involved, since a number of stimulus-design factors have to be considered, on the basis of which four hypotheses were formulated. First there is the boundary hypothesis, which states that the middle third portion of the syllable nucleus is critical to the perception of alignment distinctions. The spectral hypothesis relates discrimination to the spectral characteristics of the experimental syllable. The pitch-height hypothesis established a link between discrimination and the presentation order of the stimuli. As such it indicates that informants are more sensitive to pitch-raising then to pitch-lowering. Finally, it was pointed out that the variable of pitch movement type in the hat pattern is confounded with the placement of the alignment continuum in the stimuli. It was argued that the spectral hypotehsis and the placement of the alignment continuum in the stimuli are unlikely to have caused the differential threshold, while the boundary and pitch height hypotheses need further investigation.

In a more linguistically oriented perspective, the category-size hypothesis is formulated, which equates the JND to a proportion of the physical size of the tonal categories associated with the movements. It suggests that informants' perceptual sensitivity to pitch differences is determined by implicit knowledge about tonal categories. This interpretation of the results corroborates the tonal categorization data presented in chapter 4 and 5 and accounts for the non-uniformity of pitch perception.

2. The Differential Threshold of Pitch Movement Alignment.

In order to provide a better insight in the causes underlying the differential threshold, a follow-up study was carried out. The aims of this experiment are twofold. On the one hand, it is investigated whether the differential threshold found in the previous experiment can be replicated using a more balanced experimental design. On the other hand, the predictions derived from the boundary and pitch height hypotheses are tested. For this purpose, the same AX-format was used as in the previous experiment, but this time two additional factors were carefully controlled, the most important being the presentation order of the stimuli and the lcoation of the pitch movement in the reference stimulus.

2.1. Hypotheses and Predictions.

The *boundary hypothesis*, based on the results of the previous experiment, states that differences in pitch movement alignment can be discriminated only if the endpoints of the movements in two stimuli of an AX discrimination item are located across a critical boundary, which is assumed to be situated in the middle of the syllable nucleus. It logically derives from this hypothesis that the actual gap size between two alignment locations is not relevant. As long as both movements in an item are located on the same side of this boundary, they are not discriminated. If, however, they are located on different sides of this boundary, they should be discriminable.

This hypothesis consequently predicts that a different discrimination threshold will be found for pitch movement alignment depending on the location of the movement in the reference stimulus. In other words, the threshold of both rises and falls is expected to be equal to the gap size between the movement in the reference stimulus and the assumed boundary.

The *pitch height hypothesis* on the other hand states that alignment differences are perceived as variations in overall syllable pitch. This hypothesis predicts that the discrimination threshold depends on the presentation order of the stimuli in the discrimination items. In the rise condition, an early-late presentation order in terms of pitch movement alignment corresponds to a lowering of syllable pitch, whereas the late-early presentation order raises the pitch. Since informants have been shown to be more sensitive to pitch raising than to lowering, a large threshold is predicted for the early-late order. In the late-early order, the threshold is expected to be significantly smaller. In the fall condition, the predictions are reversed. The early-late order corresponds to raising the pitch and consequently suggests a small threshold. The late-early order lowers the pitch of the experimental syllable in a discrimination item, so that a larger discrimination threshold can be expected.

The spectral hypothesis suggests that the difference in discrimination threshold may have arisen from spectral differences between the syllables in which the pitch movements are located. Although this point is not specifically addressed in this experiment, the spectral and durational characteristics of the experimental syllables are kept constant, in order to rule out potential artefacts of these variables.

2.2. Experimental Design.

The predictions stated above are investigated by means of a series of fixed standard AX-discrimination experiments, in which subjects are presented with pairs of utterances with intonation patterns only differing in the alignment of a pitch movement. The experimental design has three independent variables: the TYPE of pitch movement in the alignment continuum (rise vs. fall), the LOCATION of the pitch movement in the standard stimulus of each discrimination item (early vs. late) and the presentation ORDER of the stimuli in the items (early-late vs. late-early).

2.2.1. Stimulus Construction.

The stimuli of this experiment are based on the Dutch utterances (3) and (4):

(3) Renaat is bij Heleen (Renaat is with Heleen).
(4) Heleen is bij Renaat (Heleen is with Renaat).

Sentence (3) was used for the stimuli in the fall-alignment task, whereas (4) served in the rise-alignment experiment. In both instances, 'leen of 'HeLEEN' was the experimental syllable on which the pitch movement continuum was implemented.

The individual constituents needed to make up these utterances were spoken by a male native speaker of Dutch and recorded in a professional studio. The constituent 'Heleen' was spoken ten times on a monotone pitch level of medium height. The carrier phrase 'Renaat is bij__' was repeated ten times with a prosodic pattern appropriate to the initial part of the hat pattern, i.e. a rise on 'NAAT', followed by a stretch of high declination. The carrier phrase '__is bij Renaat' was realized ten times with a suitable final part of the hat pattern: high declination, followed by a fall on 'NAAT' of 'Renaat'. The prosodic characteristics of these constituents are illustrated in figure 6.5:

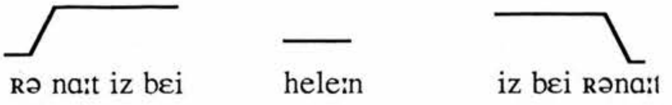


Figure 6.5: Illustration of the prosodic characteristics of the recorded constituents.

From the realizations of each constituent, one delivery was chosen which sounded fluent in all respects. These deliveries were digitized at F_S=20,000 Hz with F_C=8000 Hz. In the next stage of the stimulus construction process, the waveform of 'Heleen' was combined with those of the carrier phrases 'Renaat is bij__' and '__is bij Renaat'. This was achieved by means of a waveform recombination programme (Dryden & Mcleod, 1988). Hence, two syntactically/semantically complete and well-formed Dutch utterances were obtained, which constituted the basis of the stimuli. The advantage of this technique is that the experimental syllable 'LEEN' of 'Heleen' is identical in the two utterances, both from a spectral and durational point of view.

Subsequently, both utterances were analyzed acoustically and provided with a standard hat pattern along the same principles described in the methodological section of chapter 4. In both the rise and fall alignment conditions, the endpoint of the experimental pitch movement coincided with the onset of the vowel nucleus in 'LEEN'. The stimuli were derived by shifting the experimental movement to the right in steps of 10 msec. This yielded 42 stimuli, i.e. 21 for fall-alignment and 21 for rise-alignment. The location of the pitch movements in the experimental syllable of the stimuli is given in appendix A22 and is illustrated in figure 6.6:

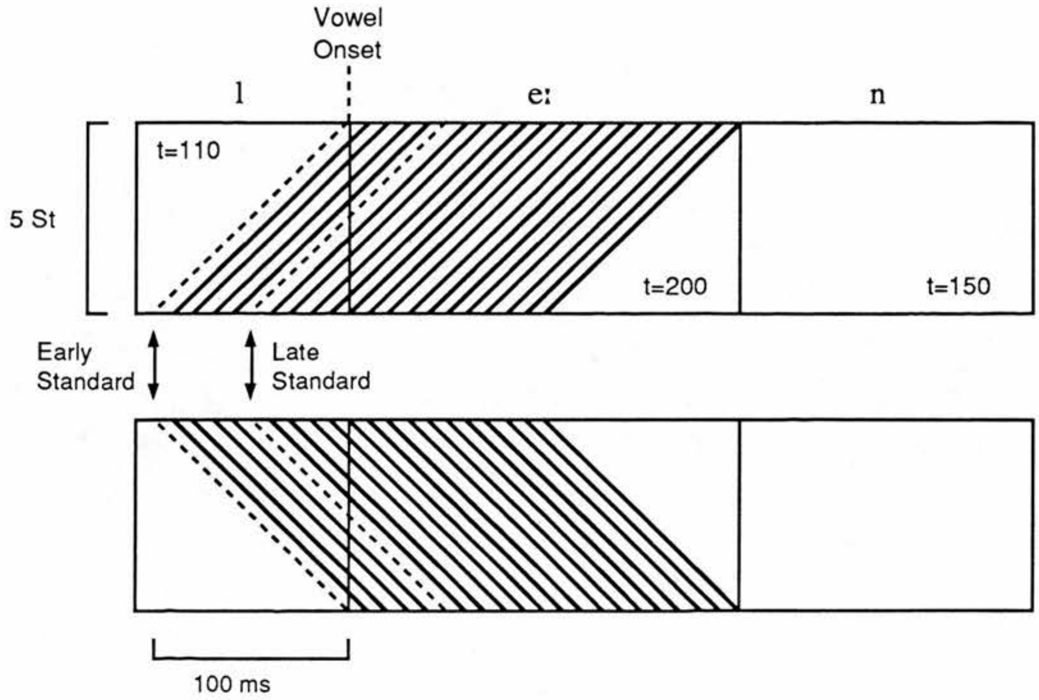


Figure 6.6: Schematization of the range of variation in the stimuli for rise alignment (top) and fall alignment (bottom). Vertical lines indicate the segment boundaries and durations are specified in milliseconds. The dashed lines represent the alignment of the early and late standards in the experimental syllable of the reference stimuli.

2.2.2. Item Construction.

For the construction of the test items, two reference stimuli were chosen. In the first reference stimulus, the pitch movement was aligned early, i.e. the endpoint of the movement coincided with the onset of the vowel nucleus of the experimental syllable. In the second reference stimulus, the endpoint of the movement was aligned 50 msec later. The alignment characteristics of the reference stimuli are found in figure 6.6. Both reference stimuli were combined into AX discrimination items with 16 other stimuli from the continuum, yielding items with gap sizes ranging from 10 to 150 msec. For each stimulus, there were two presentation orders: early-late and late-early. Each item occurred five times in the test, so that 600 discrimination items were obtained.

2.2.3. Test Tapes.

The discrimination items were sequenced in two presentation blocks. The first block contained the items for rise alignment. The second block contained those for the fall alignment task. In each block, the items were randomized across all the variables in the experimental design. Furthermore, each block was preceded by 10 trial items.

As in the former experiments, the items were not preceded by identifiers, but orientation signals were provided after every set of 10 discrimination items.

2.2.4. Informants.

11 informants took part in the experiment. They were Belgian and Dutch visiting students at the University of Aarhus (Denmark) and they took part on a voluntary basis. Neither informant had taken part in any of the former experiments.

2.2.5. Procedure.

The procedure was identical to the former experiment, except that after presentation of 150 items, the informants were given a short break in order to avoid fatigue.

2.3. Results.

In this experiment, a total of 6600 observations were obtained, i.e. 3300 for the rise and 3300 for the fall condition. The raw scores are given in appendix A23-26. The discrimination results averaged over all the experimental conditions are summarized in figure 6.7:

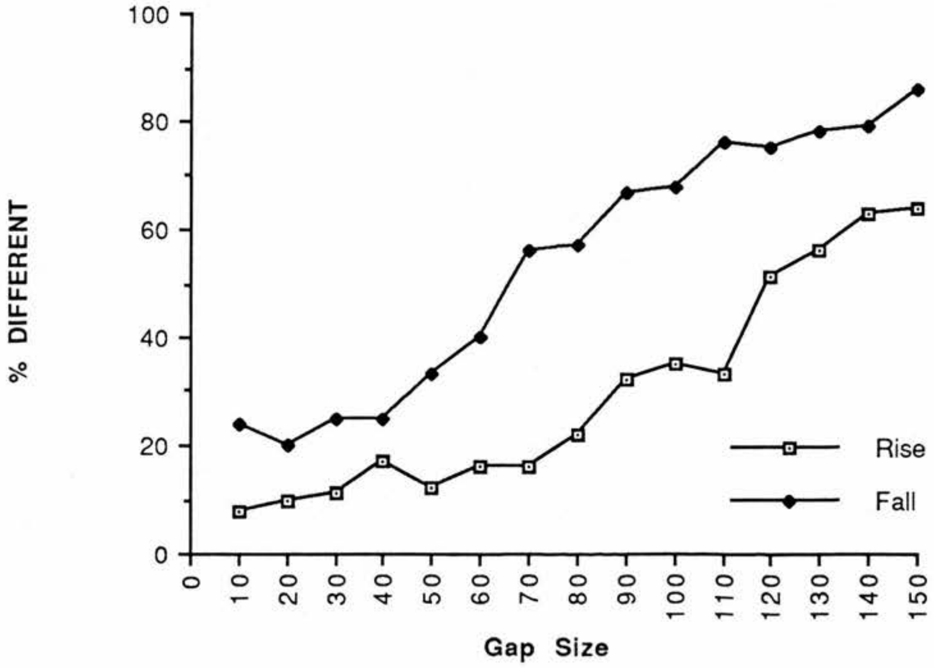


Figure 6.7: % 'different' judgements as a function of gap size (msec) for rise and fall alignment. Each datapoint summarizes 220 observations.

As in the previous experiment, the first step in the analysis consisted of establishing the JND for the individual subjects in the different experimental conditions by means of the method of average z-scores. This gives a JND of 94.86 msec for rise-alignment and 70.10 msec for fall-alignment. Hence, a difference in JND of 25 msec is observed. These values compare extremely well with those obtained in the previous experiment: the JNDs for fall-alignment are almost identical, while the JND for rise-alignment is 6 msec smaller in this experiment.

In order to investigate the relationship between the independent variables in the experimental design and discrimination performance, an analysis of variance was carried out. The independent variables in the analysis were: LOCATION of the pitch movement in the standard stimulus (early vs. late), presentation ORDER of the stimuli in the discrimination item (AB vs. BA) and TYPE of pitch movement in the continuum (rise vs. fall). The dependent variable was the JND for the individual subjects in the different experimental conditions. The results of the analysis are given in table 6.1.:

ANALYSIS OF VARIANCE

	JND	JND-measure
by	LOCATION	Rise Location in the standard
	ORDER	Presentation order
	TYPE	Type of pitch movement in continuum

Source of Variation	Sum of squares	DF	Mean Square	F	Sig.
Main Effects	13879	3	4626	6.419	.001
LOCATION	6177.8	1	6177.8	8.572	.005
ORDER	32.942	1	32.942	.046	.832
TYPE	7521.7	1	7521.7	10.436	.002

None of the higher order interactions is significant.

Explained	14504.7	7	2072.1	2.875	.015
Residual	31711.7	44	720.7		
Total	46216.5	51	906.2		

80 cases were processed

Table 6.1: Results of the analysis of variance by means of the SPSS-X software. The independent variables are LOCATION, ORDER and TYPE. The dependent variable is JND measure in the different experimental conditions.

From this analysis, it emerges that two of the main effects are statistically significant, i.e. TYPE of pitch movement in the continuum (rise vs. fall) and LOCATION of the standard in the reference stimulus (early vs. late). Furthermore, none of the higher order interactions are significant, nor is the effect of presentation order of the stimuli in the discrimination items.

The mean JNDs for the different experimental conditions are summarized in table 6.2:

ORDER	**LOCATION**	**TYPE**	
		Rise	Fall
Early-late	Early	105	81
	Late	85	60
Late-Early	Early	99	84
	Late	88	52

Table 6.2: Summary of the mean JNDs (msec) in the different experimental conditions in the design.

2.4. Discussion.

The results of this experiment provide a good replication of the results in the previous experiment, in that fall-alignment is considerably better discriminated than rise-alignment. This is clearly reflected in the JND for both types of pitch movements: 70 msec for falls and 95 for rises. These values correspond extremely well with those obtained in the previous experiment. In what follows, this result will be related to the hypotheses that were formulated in section 2.1.

On the basis of the boundary hypothesis, it was predicted that it is not gap size per se which is decisive to the discrimination of pitch movement alignment, but rather whether the pitch movements in the two stimuli of a discrimination item are located on different sides of a psycho-acoustic boundary. This boundary was postulated to be situated in the middle of the syllable's vowel nucleus. In order to find support for this hypothesis, the location of the pitch movement in the standard stimulus was manipulated as an independent variable. In one set of stimuli, the endpoint of the pitch movement coincided with the onset of the vowel nucleus of the experimental syllable. In the second set, the endpoint was located 50 msec later. The duration of the experimental syllables were identical in all stimuli. The boundary hypothesis predicted that the JND for items with an early standard is larger than the JND for items with a late standard, since it takes a bigger gap to cross the hypothetical boundary if the standard movement is located early in the syllable. Calculation of average JND in terms of location of the movement in the standard stimulus reveals that items with an early standard have a JND of 91 msec, while those with a late standard have a JND of 69 msec. Although the JND for items with an early standard is 22 msec bigger than that for items with a late standard, the difference is not big enough to compensate for the difference in alignment between early and late standard, i.e. 50 msec.

Nevertheless, it is to be recognized that the analysis of variance indicates a significant influence of the location of the standard in the reference stimulus on the size of the JND. This can be explained in terms of the spectral constraints hypothesis (House, 1990), which holds that pitch sensitivity increases as the spectral complexity of the signal decreases. If this is related to the stimulus design in this experiment, it is to be observed that the pitch movement in the reference stimulus with early standard is entirely situated in the prevocalic consonant, so that the alignment differences are situated in the spectrally loaded portion of the experimental syllable around the segment boundaries between [l] and [e:]. In the reference stimulus with a late standard, the alignment differences are located in a spectrally more stable portion. It is not unlikely that the difference in JND obtained for items with early and late standard is related to the relationship between spectral processing load and pitch discrimination proposed in House (1990), since pitch sensitivity increases in spectrally stable portions of the signal.

As a final point regarding the location of the standard in the reference stimulus, it has to be mentioned that this variable may have made a small contribution to the differential threshold in the previous experiment, since the standard in the reference stimulus for fall alignment was located later in the experimental syllable nucleus than in the reference stimulus for rise alignment. As a result, this variable was confounded with the type of pitch movement. This may account for the fact that the difference in JND between rise and fall alignment in the previous experiment was somewhat bigger than the one observed in this experiment, i.e. 33 msec vs. 25 msec.

The second hypothesis stated that informants perceive pitch movement alignment in terms of variations in overall pitch level of the experimental syllable and it is expected that subjects are more sensitive to pitch-raising than to a lowering of pitch. If these two observations are related to the presentation order of the stimuli in the discrimination items, it is predicted that informants are most sensitive to the BA presentation order in the rise-alignment task. In the fall-alignment task, a greater sensitivity to the stimuli in AB order is to be expected. These respective orders involve a raising of the overall pitch level of the syllable, whereas the reverse orders involve a lowering of the syllable pitch. In order to examine whether these predictions are borne out, a JND value was calculated for both presentation orders in the two pitch movement continua. In the rise condition, the JND AB is 105 msec, while the JND BA is 85 msec. In the fall condition, the JND AB is 81 msec and the JND BA is 60 msec. From this, it can be concluded that the predictions of the pitch height hypothesis are not substantiated, since the JNDs for both types of pitch movements are larger in AB presentation order than in BA order.

2.5. Conclusions.

The discrimination thresholds for rise and fall alignment in the hat pattern obtained in this experiment agree very well with the values established in the previous experiment. The most important observation regards the nature of these thresholds, in that informants' differential sensitivity to rise and fall alignment is clearly substantiated. As a result of the balanced experimental design, several causes that were assumed to underlie this differential threshold can be ruled out. The results show that it cannot be accounted for in terms of durational and spectral differences between the experimental syllables in both conditions since these variables were kept constant. Furthermore, the boundary and the pitch height hypotheses could be eliminated with a high degree of confidence. In addition, it was argued earlier that the confounding effect, if any, between the type of pitch movement and the overall placement of the alignment continua in the stimuli can be regarded as minimal. As a result, the interpretation of the threshold difference that gradually emerges is one in which it represents a differential response to rises and falls, or one in which the difference is accounted for in terms of the fact that the movements occur in a specific prosodic constext, i.e. the hat pattern.

3. Alignment of Isolated Rises and Falls.

In order to provide evidence in favour of the above-mentioned interpretations, it was decided to carry out an additional experiment in which the JND of pitch movement alignment is investigated for rising and falling pitch movements occurring in isolation. The hypotheses and predications can be formulated as follows. If the observed differential threshold from the previous experiments represents a differential response to rising and falling pitch movements, irrespective of the pattern in which they occur, it is to be expected that a similar differential sensitivity is observed for alignment differences in isolated rises and falls, i.e. movements which do not occur as part of the hat pattern. If on the other hand no significant difference is found between the JND's of rise and fall alignment of isolated movements, it is shown indirectly that the differential threshold of the previous experiments is to be related to the fact that the movements were embedded in the hat pattern.

In order to investigate these predictions, two fixed standard AX discrimination experiments were carried out in which informants were asked whether they could discriminate between pairs of stimuli which differed from each other in terms of the position of a pitch movement in an accented syllable. The independent variables in the experimental design are the TYPE of pitch movement (rise vs. fall), presentation ORDER (AB vs. BA), GAP size between the locations of the movements in the stimuli (15-150 ms) and the LOCATION of the movement in the reference stimulus (early vs. late).

3.1. Experimental Design.

3.1.1. Stimuli.

The stimuli for this experiment were prepared by resynthesizing modified natural speech, the method of which was described previously. For this experiment, the utterance [hele:na] was chosen as the basis for the stimuli. In this utterance, all the segments are voiced and [le:] carries the primary accent: it is preceded and followed by a single unstressed syllable. The stressed syllable was used as the experimental syllable with which pitch movement alignment distinctions are associated. The utterance was read 10 times by a male native speaker of Dutch on a monotone pitch of medium height. The utterances were recorded in a sound-treated room and from these recordings, one delivery was chosen which sounded fluent in all respects. The utterance was digitized at F_S=20.000 Hz/F_C=8000 Hz and analyzed acoustically by means of the API subroutine of the ILS signal processing package. On this utterance, four pitch movement alignment continua were implemented.

In order to obtain the stimuli for two rise-alignment continua (RA), the utterance was given a pitch contour in which a rising pitch movement was positioned in the accented syllable. The rise was preceded by a stretch of low declination on the first syllable and followed by high declination extending into the third syllable. Duration of the rise was 100 ms with an excursion size of 5 semi-tones. In the first continuum, the endpoint of the rise was located 5 ms after vowel onset. This stimulus will be referred to as the reference stimulus with the early standard. From this contour, 11 stimuli were derived by shifting the rise to the right in steps of 15 msec, leaving the remaining part of the contour unchanged. In the second rise-alignment continuum, the endpoint of the rise in the reference stimulus was located 105 msec after vowel onset (late standard) and 11 stimuli were derived by shifting the rise to the right in steps of 15 msec. The alignment of the rises to the segmental string in both continua is illustrated in figure 6.8:

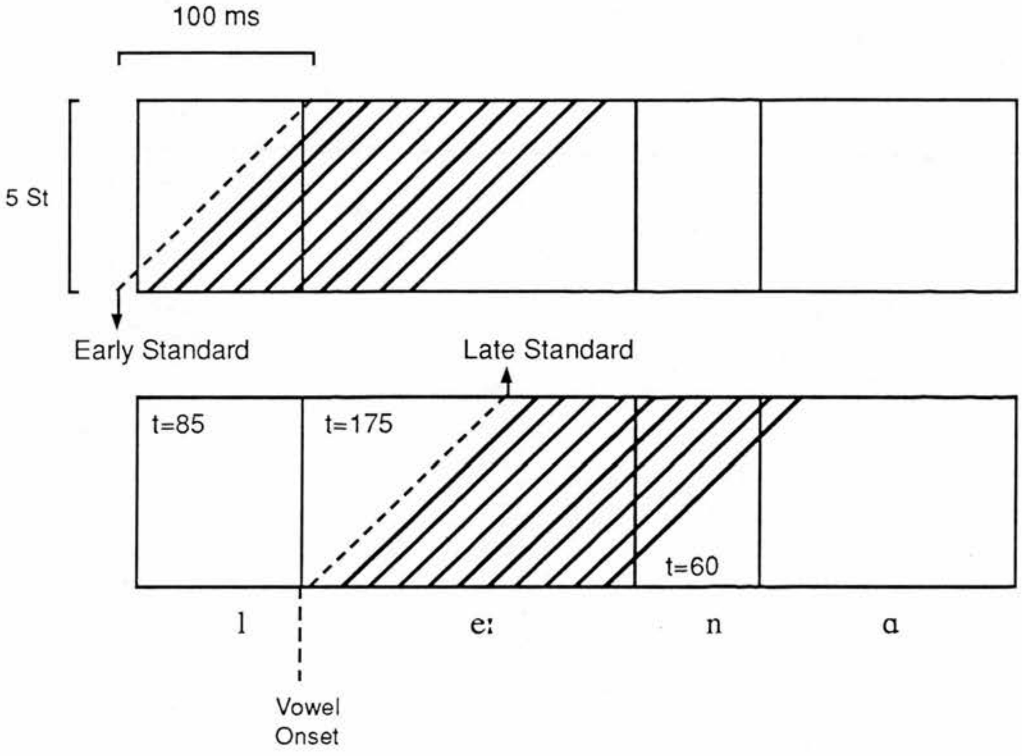


Figure 6.8 : Location of the rising pitch movements relative to the segmental string in the stimuli for the two rise-alignment continua. Vertical lines indicate segment boundaries and segment durations are specified in milliseconds. Dashed lines indicate the alignment of the standard pitch movements to the experimental syllable in the reference stimuli.

The stimuli for the fall-alignment continua (FA) were derived from the same realization of the utterance as the stimuli in RA, so that the segmental durations and spectral characteristics were identical. The utterance was given a pitch contour consisting of a falling pitch movement in the accented syllable, preceded by a stretch of high declination and followed by low declination. Duration, excursion size and location of the endpoint were identical to those of the rises. From this reference contour with early standard, 11 stimuli were derived by shifting the fall to the right in steps of 15 ms. The same procedure was applied to obtain 11 stimuli for a second fall alignment continuum, in which the fall endpoint in the reference stimulus was located 105 ms after vowel onset (late standard). The alignment of the falls in both continua is illustrated in figure 6.9:

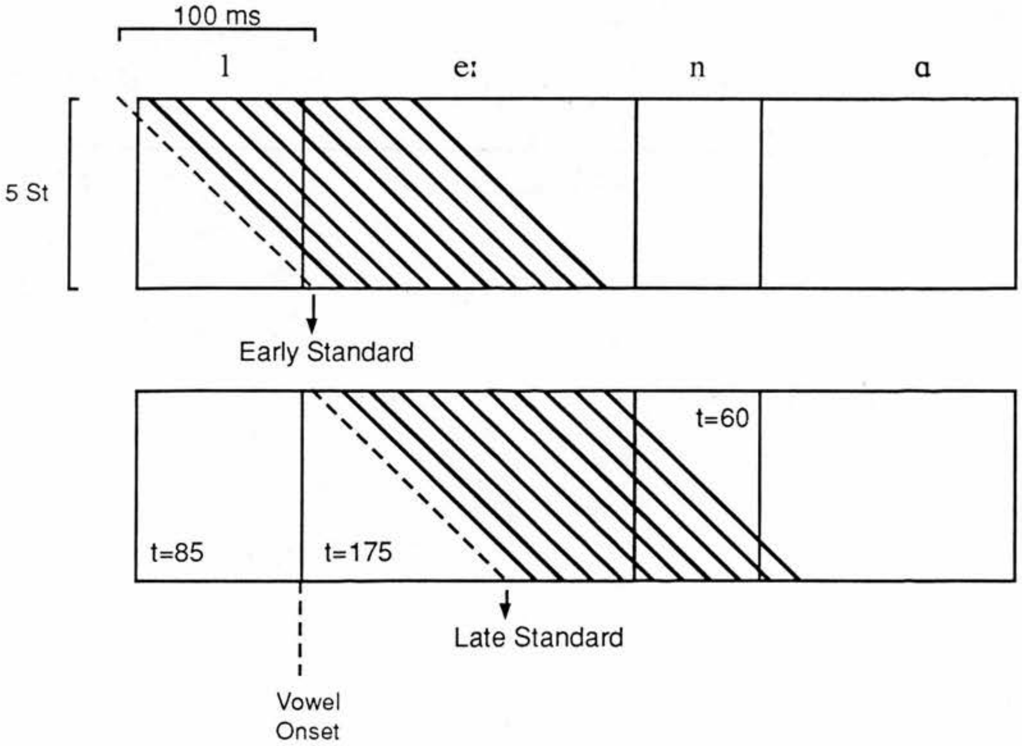


Figure 6.9 : Location of the falling pitch movements relative to the segmental string in the stimuli for the two fall-alignment continua. Segment durations are specified in milliseconds. The dashed lines indicate the alignment of the standard pitch movements in the experimental syllable of the reference stimuli.

3.1.2. Items.

In order to obtain the discrimination items, the reference stimuli from each alignment continuum were combined with every other stimulus in the continuum as a result of which alignment differences between reference stimulus and comparison stimulus ranging from 15-150 ms were obtained. In order to counterbalance for order effects (Woodworth & Schlosberg, 1954), two presentation orders were used. In AB order, the stimulus with the earlier movement came first. In BA order, the stimulus with the later movement was presented first. The items in each experimental condition occurred 4 times in the test, so that a total of 320 discrimination items were obtained, i.e. 160 for RA (10 items x 2 standards x 2 orders x 4 repititions) and 160 for FA.

In order to reduce informants' listening time, AA and BB items were omitted from the test.

3.1.3. Test Tapes.

Two test tapes were recorded, each of which consisted of two presentation blocks. On the first tape, the first block contained the discrimination items for RA, randomized across all the variables in the experimental design, i.e. location of the standard in the reference stimulus (early, late), gap size (15-150 ms) and presentation order (AB, BA). The second block consisted of the items for FA with a different randomization. On the second tape, the first presentation block contained the items for FA. The randomization of these items was identical to the one for RA on tape 1. The second block consisted of the items for RA, the randomization of which was identical to the items for FA on tape 1.

The inter-stimulus interval in the items was 500 msec, while the inter-item interval amounted to 2500 msec. There were no identifiers on the tape, but after each set of 10 items, the informants heard an orientation signal.

3.1.4. Informants.

Two groups of informants were recruited (N=28) from the student population of the Department of Germanic Philology and the Department of Business Studies of the University of Antwerp (UFSIA), Belgium. They were all native speakers of Dutch and took part on a voluntary basis. 15 informants listened to tape 1 and 13 to tape 2. Each informant participated in only one experimental session.

3.1.5 Procedure.

The informants were seated in a quiet language laboratory, in which the volume level of the headphones was individually adjustable. They were told that they were going to take part in an experiment on Dutch intonation, which aimed to establish the intonational differences that are perceivable by native speakers of Dutch. Informants were instructed to report on whether the speech melody on the experimental syllable of each wordpair was *same* or *different*.

Before the start of the experiment, the subjects heard 10 trial items to get used to the task and adjust the volume to a comfortable listening level. Subsequently, they proceeded to the two presentation blocks in each experimental session. Informants were given a 5 minute break between the two presentation blocks. Total duration of each tape was 32 minutes.

3.2. Results.

In this experiment, a total number of 10.192 observations were obtained, i.e. 5094 for RA and 5094 for FA. The discrimination judgements are summarized in figure 6.10:

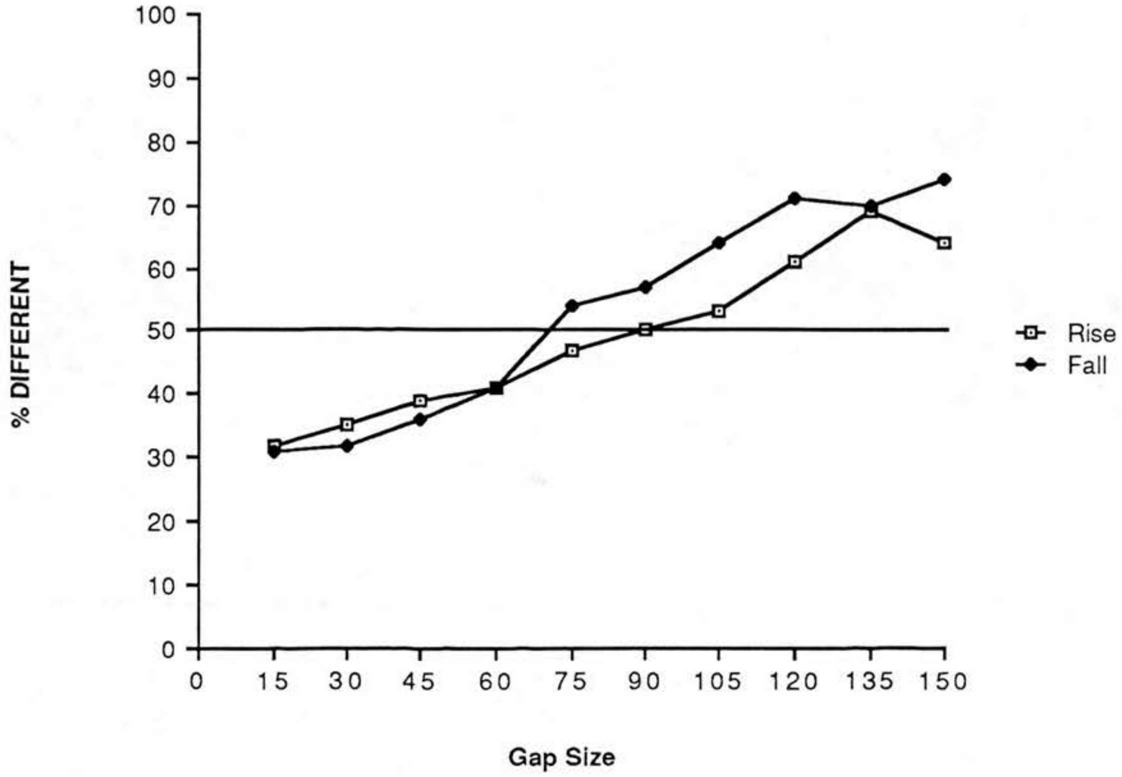


Figure 6.10 : % 'Different' judgements as a function of gap size. Each datapoint summarizes 112 observations.

In the first instance, the JNDs for pitch movement alignment in the different experimental conditions were calculated for the individual subjects. This was achieved by the method of average z-scores as explained in section 1.2 of this chapter. The JNDs established by this procedure were subsequently submitted to an analysis of variance. The independent variables in this analysis were the TYPE of pitch movement (rise vs. fall), the LOCATION of the standard in the experimental syllable of the reference stimuli (early vs. late) and the presentation ORDER of the stimuli in the discrimination items (Early-late vs. late-early). The result of this analysis is given in table 6.3:

ANALYSIS OF VARIANCE

	JND	JND-measure
by	LOCATION	Rise Location in the standard
	ORDER	Presentation order
	TYPE	Type of pitch movement in continuum

Source of Variation	Sum of squares	DF	Mean Square	F	Sig.
Main Effects	11257	3	3752.6	3.543	.016
LOCATION	6564.6	1	6564.6	6.199	.014
ORDER	5127.4	1	5127.4	4.842	.029
TYPE	157.4	1	157.4	.149	.700

None of the higher order interactions is significant.

Source of Variation	Sum of squares	DF	Mean Square	F	Sig.
Explained	20327.02	7	2903.9	2.742	.010
Residual	162031.1	153	1059.02		
Total	182358.12	160	1139.74		

224 cases were processed

Table 6.3: Results of the analysis of variance by means of the SPSS-X software. The independent variables are LOCATION, ORDER and TYPE. The dependent variable is JND measure in the different experimental conditions.

From table 6.3, it emerges that the main effects of LOCATION and presentation ORDER are statistically significant, whereas the TYPE of pitch movement is not. As far as the higher order interactions are concerned, the only significant interaction is one between TYPE and LOCATION. The mean JNDs in the different experimental conditions are summarized in table 6.4:

ORDER	**LOCATION**	**TYPE**	
		Rise	Fall
Early-late	Early	92	101
	Late	75	83
Late-Early	Early	94	70
	Late	77	72

Table 6.4: Summary of the mean JNDs in the different experimental conditions of the experiment.

3.3. Discussion.

The most important conclusion that can be arrived at on the basis of the data in this experiment is that there is no significant difference in the JNDs of pitch movement alignment as a function of the TYPE of pitch movement in the continuum: the JND for rise alignment is found to be 84 msec, whereas this for fall alignment is 82 msec. This can be taken as a clear indication of the fact that the differential threshold found in the previous experiments does not represent a general differential response to the alignment of rises and falls. If falls were generally better discriminated than rises, smaller JND is to be expected for fall alignment than for rise alignment. It is clear than any such difference does not materialize.

This strongly suggests that the differential threshold in the previous experiments is to be accounted for by the fact that the movements in the stimuli constitute an integral part of the hat pattern. Although informants participating in these experiments were specifically instructed to listen to any melodic differences in the experimental syllable only, it can be assumed that they hear the pattern as a whole and recognize it as an instance of the hat pattern. On the basis of the categorization experiment reported in chapters 4 and 5, it can be assumed that the exact position of the rise in the experimental syllable is of comparatively little importance to the identity of the pattern. The location of the fall, however, is of greater perceptual importance, since alignment distinctions can be exploited to distinguish the hat pattern from the so-called cap pattern. In the former, the fall is located late in the syllable, whereas in the latter, the fall is located very late.

The interpretation of the differential threshold proposed here suggests that discrimination of pitch features is to some extent determined by linguistic knowledge, in that it is more accurate with respect to features which can be exploited for linguistic purposes, than with respect to features which cannot bring about linguistic distinctions. Such an interpretation is not new, since this phenomenon has been demonstrated with respect to a large number of segmental aspects of speech in categorical perception research. What is interesting about this observation is that the same phonetic dimension of pitch movement alignment is perceived differently depending on the prosodic context in which the differences occur.

If this interpretation is correct, it has important implications for the analysis methodology of intonation, in that observations on the discrimination and perception of pitch features obtained in a particular prosodic context cannot be readily generalized to other prosodic contexts. This is precisely what has been done in for instance the Grammar of Dutch Intonation, in which the categorization of perceptually relevant pitch movements is based entirely on experimentation into the tolerance levels of native speakers of Dutch with regard to various pitch features. It was shown in chapter 3 that the knowledge about the perception of these pitch features obtained in a particular prosodic context is generalized to a large variety of other prosodic contexts to standardize the pitch movements in these intonation patterns. The results presented in this chapter strongly indicate that this procedure is questionable, so that is can be expected that a number of generalizations proposed in the GDI may have to be revised.

Concerning the other variables in the experimental design, the analysis of variance indicated a significant effect of the presentation ORDER of the stimuli in the discrimination items and the LOCATION of the standard in the reference stimulus. In addition, a significant interaction between TYPE of pitch movement and presentation ORDER was obtained. The interaction indicates that the JND for rise alignment in early-late order (81 msec) is smaller than the JND in the late-early order of presentation (86 msec). For fall alignment, the opposite is observed: the JND in early-late order (93 msec) is larger than in the late-early presentation order (71 msec). These observations are precisely the opposite of what was predicted on the basis of the pitch height hypothesis in section 2.1 of this chapter, and suggest that informants are more sensitive to pitch lowering than to pitch raising. If indeed the alignment continua are perceived in terms of pitch height, the early-late presentation order in items for rise alignment lower the overall pitch of the experimental syllable, while a similare effect is achieved in the late-early order in items for fall alignment. This assymetry can consequently account for the significant interaction if it is assumed that informants are most sensitive to pitch lowering. This observation is inconsistent with 'the results of informal experimentation reported earlier, which suggested that informants are more sensitive to pitch raising. This apparent inconsistency cannot readily be explained, but suggests that also along the dimension of perceived pitch height the non-uniformity principle has to be taken into account.

The location of the standard in the reference stimuli finally, was also observed to have a significant influence on the size of the JND, in that a somewhat smaller JND is found for items with a late standard in the reference stimulus: 89 msec vs. 77 msec. This is entirely in agreement with the observations in the previous experiment, in which a similar difference was found. This is consistent with the spectral constraints hypothesis that alignment differences are more difficult to perceive in spectrally loaded portions of the speech signal: in the stimuli with early standard, the standard movement is located entirely in the prevocalic consonant, whereas in those with late standard, the standard pitch movement was located entirely in the spectrally steady-state part of the vowel nucleus in the experimental syllable (cfr. figures 6.8 and 6.9).

3.4. Conclusions.

The results of this experiment on the just noticeable difference of pitch movement alignment indicate that there is no differential threshold as a function of the TYPE of pitch movement. This suggests that the differential sensitivity of informants observed in the previous experiments does not represent a general differential response to rises and falls. This is taken as evidence for the fact that discrimination is to some extent determined by implicit linguistic knowledge. It also indicates that the uniformity hypothesis cannot be maintained. The implications of this for the GDI were briefly discussed.

The significant interaction between type of pitch movement and the presentation order of the stimuli in the discrimination items suggest that informants are most sensitive to a presentation order which involves lowering overall syllable pitch. The fact that this is not consistent with the experiments on alignment discrimination in the hat pattern and informal experimentation on the perception of peak height is difficult to account for. In any case, it provides additional counterevidence to the uniformity hypothesis.

Finally, the differential threshold as a function of the location of the standard in the reference stimulus is interpreted as consistent with the spectral constraints hypothesis.

4. Rise Alignment Categorization: Further Evidence.

A last point which remains to be addressed in this chapter is the fact that the data presented in the earlier sections of this chapter are not entirely unproblematic for the categorization of rise-alignment. The JND that is obtained here is considerably larger than what could have been expected on the basis of information in the literature. The consequence of this is that the conclusions regarding the categorization of rise-alignment in chapter 4 may be questioned. The gap sizes used in the set-up of that experiment ranged between 30 and 70 msec, which is substantially smaller than the JND for rise-alignment. Consequently, the absence of any discrimination peaks in the discrimination curve may just be an indication of the fact that it is psycho-acoustically impossible to hear any differences between the stimuli, so that the discrimination peak could not materialize. Therefore, it was decided to carry out an additional experiment into the categorization of rise-alignment in the hat pattern.

In order to re-examine the categorization of rise-alignment in the hat pattern, a 2IAX-discrimination experiment was carried out, the design of which was similar to those reported in chapters 4 and 5. For this purpose, a rise-alignment continuum was prepared in the same fashion as in chapter 4. The sentence that constituted the basis for the stimuli was the sentence 'Renaat is ziek'. The rise in the hat pattern had an excursion size of 5 semi-tones with a duration of 100 msec.

The stimuli were combined into variable standard AX-discrimination items in such a way that the gap size between the endpoints of the rises in both stimuli was 100 msec, a value which corresponds to the JND. This yielded 26 discrimination items. The stimuli were presented in both AB and BA presentation order and each item occurred 5 times in the test so that 260 items were obtained (26 x 2 x 5).

The items were presented to 10 students of the 'Rijkshogeschool voor Vertalers en Tolken' (Brussels) who had not taken part in any of the previous experiments. They were all native speakers of Dutch and they participated on a voluntary basis. The experimental procedure was identical to that described in chapter 4.

In this experiment, a total number of 2,600 observations were obtained. The raw scores are given in Appendix A27. The discrimination data are summarized in figure 6.11:

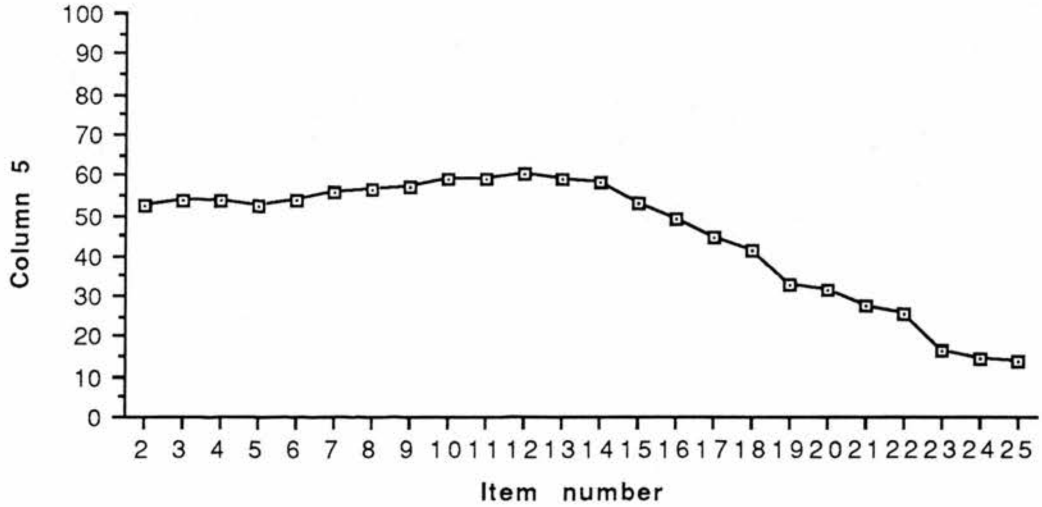


Figure 6.11: % 'Different' judgements as a function of item number. The gap size between the rises is 110 msec. Each data point represents 50 observations.

Figure 6.11 reveals that average items discrimination is quite substantially better than in the rise-alignment experiment reported in chapter 4. However, it remains near chance level, except for the items at the far right of the continuum, where discrimination decreases gradually. More importantly, the curve clearly has no discrimination peaks. This is entirely consistent with the rise discrimination data presented in chapter 4 and thus supports the conclusion that rise-alignment in the Dutch hat pattern is not perceived in terms of underlying tonal categories, but constitutes a single perceptual category. Since the gap size between the stimuli of the discrimination items in this experiment is above the JND, the flatness of the curve cannot have been caused by the fact that the physical differences in rise-location were too small from a psycho-acoustic point of view.

5. Conclusions.

In this chapter, the discrimination threshold of pitch movement alignment was investigated. The most important conclusion of these experiments is that the JND difference of rise-alignment is substantially larger than that of fall-alignment. Several psycho-acoustic and linguistic hypotheses were proposed to account for this difference. Further experiments have shown that the predictions of the boundary hypothesis, the pitch height hypothesis and the spectral hypothesis cannot be substantiated. It was also shown that the differential threshold is not the result of a differential perceptual response to rises and falls. Consequently, a more linguistically oriented explanation in terms of the phonological status of the accents associated with the pitch movements or the size of the perceptual categories may have to be considered. More experimental work in this area, however, has to be carried out.

Finally, a follow-up experiment on rise-discrimination was reported, in which the gap size between the rise locations was increased to a value above the JND. The results of this experiment are entirely in agreement with the those of chapter 4, in that there is no discrimination peak. This is taken as indicative of the fact that rising pitch movements are associated with a single underlying category of alignment in the hat pattern.

CHAPTER 7

General Conclusions.

0. Introduction.

In this chapter, the most important conclusions of the experimental work reported in this dissertation are briefly summarized and some directions for future research are indicated.

Concerning the parameter of pitch movement alignment in the Dutch hat pattern, the experimental work reported in this dissertation addresses two fundamental issues. In the first instance, the sensitivity of informants to differences in pitch movement alignment is investigated. Secondly, the categorization of alignment continua is examined. In the next sections, the results regarding each of these research topics are summarized.

1. The Perception of Pitch Movement Alignment.

As far as perception of differently aligned rising pitch movements in the hat pattern is concerned (chapter 4), it is observed that the average discrimination of the items is extremely low, i.e. 36%. From this, it is concluded that the informants are not able to discriminate between the stimuli in the items, although the actual physical differences between the stimuli conform to the traditionally accepted values suggested in Hill & Reid (1977) and Collier (1983). This conclusion is not only supported by low average discrimination, but also by a detailed analysis of subject response behaviour: the response coincidence analysis indicates that the majority of informants (60%) can be classified as non-discriminators. The average discrimination of subjects in this group amounts to 22%. The remaining informants can be assigned to the discriminator group, which distinguishes the items in 50% of the cases.

The alignment of falling pitch movements in the hat pattern (chapter 5) is better discriminated. The fact that there are significantly more discriminators (79%) than non-discriminators (21%) is indicative of this. In addition, it is observed that average item discrimination for fall-alignment is, on the whole, better than for rise-alignment. This is remarkable when it is taken into account that 80% of the physical differences between the stimuli in the items for rise alignment are equal or bigger than those for fall alignment. In the rise-alignment task, the gap size between the rises in the stimuli ranges between 30 and 70 msec, whereas in the fall-alignment task the informants are confronted with gap sizes of 40 and 50 msec. The observation that fall-alignment is better discriminated than rise-alignment is further investigated by establishing the just noticeable difference of pitch movement alignment (chapter 6). In the first instance, a fixed standard 2IAX discrimination experiment is reported in which informants were asked to evaluate items with increasing physical differences in gap size. The JND for rise-alignment is found to be 107 msec, whereas this for fall-alignment is 66 msec. This result does not only indicate that subjects are more sensitive to fall-alignment, but also that the JND for rise-alignment is considerably larger than what could have been expected on the basis of previous reports in Collier (1983). In order to explain this differential sensitivity, several hypotheses are formulated which account for this phenomenon in terms of psycho-acoustic and/or linguistic factors.

In an acoustic perspective, the boundary hypothesis was formulated. This hypothesis accounts for the differential sensitivity to pitch movement alignment in terms of a hypothetical psycho-acoustic boundary in the middle of the vowel nucleus of the syllable with which the pitch movement is associated. It states that differently aligned pitch movements are discriminated only if they are located on different sides of this boundary. The psycho-acoustic boundary is hypothesized to be located in the middle of the syllable nucleus. It is observed that this hypothesis accounts for a number of differences in the stimulus design in both alignment conditions, esp. regarding the different location of the standard pitch movement and the unequal duration of the vowel nucleus in both types of stimuli. In addition, the boundary hypothesis is consistent with research reported in Rossi (1971, 1978), who showed that the middle portion of the syllable nucleus is critical to the perception of the pitch level of F0 transitions. This hypothesis predicts that alignment differences as such are not critical to discrimination, but rather that pitch movements are discriminated provided they are located on different sides of the psycho-acoustic boundary. The hypothesis suggests that physical differences in alignment can be discriminated which are much smaller than the just noticeable differences proposed above.

An alternative explanation for the differential threshold is suggested by relating differences in pitch movement alignment to perceived pitch height (pitch-height hypothesis). This hypothesis originates from the observation that the discrimination items of the two alignment continua are asymmetrical in terms of the presentation order of the stimuli (early-late), whereas they are symmetrical in terms of the pitch height of the experimental syllables in the stimuli. More concretely, the early-late presentation order of the stimuli in the items for rise-alignment corresponds to a lowering of the pitch level of the experimental syllable. In the items for fall-alignment, the same presentation order of the stimuli corresponds to a raising of the pitch level. This asymmetry can subsequently be related to the results of other experiments, which indicate that subjects are more sensitive to pitch raising than to a lowering of the overall pitch level. Although the results of this experiment cannot directly be compared to other experimentation in this area, it should be pointed out that the pitch-height hypothesis is highly consistent with results reported in Nabelek et al. (1970), which suggest that informants indeed perceive pitch transitions in terms of overall pitch levels on syllables. It is also consistent with the findings of 't Hart (1981), who found that native speakers of Dutch are most sensitive to excursion size variations in rises.

The final hypothesis concerning the differential threshold is formulated in terms of the spectral characteristics of the experimental syllables in the alignment continua. In the rise-alignment continuum, the experimental syllable contains a monophthong, whereas the fall-alignment continuum is implemented on a diphthong, which is characterized by continuous spectral change. Collier (1983) indicates that informants are more sensitive to pitch features in areas of spectral change and this can account for the greater sensitivity to fall-alignment.

The above-mentioned psycho-acoustic hypotheses essentially originate from a set of stimulus design factors, which cause the experimental conditions for rise-alignment to be slightly different from those for fall-alignment. If abstraction is made from these factors, an interpretation of the differential threshold in terms of linguistic structure can be attempted, in that it has to be observed that the two alignment continua correlate with different linguistic categories of prominence. In the rise-alignment continuum, the rises are associated with a prenuclear accent, whereas the falls realize a nuclear accent on the utterance-final syllable. It is not inconceivable that the differential threshold is the manifestation of this linguistic difference, since it may well be the case that informants are more sensitive to phonetic variations relating to the linguistically dominant category.

The second linguistic interpretation relates the discrimination threshold to the physical size of the tonal category underlying both continua. Experimentation on the categorization of rise-alignment in the hat pattern (chapter 4) indicates that all the rises located in the vowel nucleus and the postvocalic consonant constitute a single tonal category. The categorization experiment on fall-alignment (chapter 5) suggests that the fall continuum relates to two underlying tonal categories. The falls which are entirely located in the syllable nucleus constitute the 'early' category, whereas those with the endpoint in the postvocalic consonant can be regarded as instances of the 'late' category. The result of this is that the same stretch of speech at the segmental level relates to a different number of underlying tonal categories, with the result that the physical size of the alignment categories is different. In this perspective, it is possible to take the duration of the segments as an indication of the physical size of the underlying tonal categories: the category of rises is equal to the duration of the vowel nucleus and the postvocalic consonant. The category of early falls is equal to the duration of the vowel nucleus only, while the category of late falls is equal to the duration of the postvocalic consonant. If the obtained just noticeable differences are expressed as proportions of the size of these categories, the JND is virtually identical in rises and falls, i.e. one third of the size of the tonal category. This hypothesis is consistent with the categorization data for both types of pitch movements, and indicates that the sensitivity of the human perceptual system to pitch features is mediated by implicit knowledge about intonational structure.

In order to get a better understanding of the contribution of above-mentioned psycho-acoustic processes to the differential threshold in pitch movement alignment, a follow-up study was carried out. This experiment was specifically designed to test the predictions of the psycho-acoustic hypotheses. For this purpose, the spectral characteristics of the experimental syllables, the presentation order of the stimuli in the discrimination items and the location of the pitch movement in the standard stimulus were carefully controlled. The JNDs established in this experiment are entirely in agreement with the previously obtained values: the JND for rise-alignment amounts to 126 msec while that of fall-alignment is 66 msec. A detailed analysis of the relationship between the location of the pitch movement in the standard stimulus and the just noticeable difference reveals that the JND for items with an early standard is 103 msec, while that for items with a late standard amounts to 84 msec. Although this difference is in agreement with the predictions of the boundary hypothesis, it is not statistically significant. Moreover, the 19 msec in JND does not compensate for the 50 msec difference in the location of the standard pitch movements in the early and late conditions. On the basis of these results, the boundary hypothesis cannot be maintained as an explanation of the differential threshold.

As far as the relationship between presentation order and the value of the just noticeable difference is concerned, the pitch-height hypothesis predicts that the JND for rise- alignment is smallest in the BA order, since this order corresponds to a raising of the pitch level of the experimental syllable. The JND for fall-alignment is expected to be smallest in the AB presentation order. The results of the experiment show that these predictions are not borne out: in both conditions, the JND is smallest in the AB presentation order and the effect of presentation order generally is not statistically significant. This suggests that the pitch-height hypothesis has to be rejected.

Finally, it has to be pointed out that the spectral hypothesis cannot account for the differential threshold in this experiment. Unlike the previous experiment, the experimental design here is such that the spectral characteristics of the experimental syllable are identical in both alignment conditions.

Although this experiment was not specifically designed to evaluate the category-size hypothesis, the obtained data were analyzed in this perspective. For this purpose, the JND was normalized in terms of the category size associated with both types of pitch movements. It can be concluded from this normalization process that the proportional JNDs are almost identical for both rise- and fall-alignment. This confirms the expectation that the JND has to be regarded as an invariant proportion of the tonal category size. As to the value of this proportion, a gap size corresponding to one third of the tonal category is proposed.

From all these observations, it can be concluded that a purely psycho-acoustic explanation of the differential discrimination threshold is very difficult to maintain. It is of course possible that the difference in JND is just the reflection of an intrinsic differential perceptual response to rising and falling pitch movements. This was investigated in a separate experiment, in which the JND of alignment was investigated with respect to isolated rises and falls. The results indicate that there is no significant difference in JND, from which it can be concluded that the perceptual response to rises and falls is essentially identical. This suggests that a linguistically oriented interpretation of the differential threshold has to be considered.

The hypothesis that the differential threshold of alignment discrimination originates from a differential perceptual response to rises and falls can be investigated by establishing the just noticeable difference of rise- and fall-alignment in an isolated context, i.e. in a contour which consists of a rise or a fall only. For this purpose, a pitch movement continuum can be implemented on the experimental syllable 'heLEEN', so that the experimental conditions are very similar to these of the experiments reported in chapter 6. If the results of this experiment would reveal a similar differential threshold of pitch movement alignment, a psycho-acoustic explanation in terms of the type of pitch movement involved would have to be reconsidered. Work along these lines is presently being carried out.

Even though the category-size hypothesis is substantially more speculative than the psycho-acoustic explanations, it deserves very careful consideration since it suggests that pitch discrimination in speech is mediated by the listener's implicit knowledge about intonational categories. More systematic work in this area may prove fruitful and can be attempted by establishing the JND for both rises and falls as a function of the segmental durations of the experimental syllables. With the techniques available, it is possible to vary the duration of the syllable nucleus systematically and implement a pitch movement continuum on these syllables with different durations. In order to confirm the category-size hypothesis, the results would have to show a differential threshold for pitch movement alignment which can be normalized as a function of category size. In addition, a decrease in the value of the JND would have to be observed as the duration of the vowel nucleus decreases.

2. The Categorization of Pitch Movement Alignment.

The categorization of pitch movement alignment in the Dutch hat pattern was investigated in chapters 4 and 5. In chapter 4, rising pitch movements in the hat pattern were investigated by means of a traditional 2IAX discrimination experiment and an identification task in which informants were asked to label the stimuli in terms of the psycho-acoustic labels 'high' and 'low'. A similar set-up was used to examine the categorization of fall-alignment in chapter 5.

In experimentation of this kind on segmental aspects of speech, informants' categorization is reflected by a significant peak in the discrimination curve of the discrimination task. This peak indicates that subjects respond differentially to equal physical differences in the dimension under investigation. If this is applied to rise-alignment, it is to be expected that physical differences in rise-alignment are discriminated in some areas of the continuum, whereas in other areas they are not discriminated. Such a differential discrimination behaviour can be accounted for by assuming that the continuum represents two categories. Subjects do not discriminate, if the stimuli are taken from the same category, whereas they do discriminate stimuli representing different categories. As far as the labelling results are concerned, a sudden cross-over in category identification is to be expected, which coincides with the location of the discrimination peak.

An evaluation of the data obtained for rise-alignment in chapter 4 against these criteria clearly indicates that none of these characteristics materialize. In the first instance, it is observed that there are no consistent discrimination peaks in the discrimination curves based on the raw discrimination data. In addition, a re-analysis of the discrimination data in terms of the mean rise position of the experimental items yields an essentially flat discrimination curve. This flat curve is characteristic for the data considered in terms of both rise endpoint and rise onset. It was pointed out in chapter 4, and it is emphasized again here, that the flatness of this curve is most unlikely to result from the averaging procedures used: Hill & Reid (1977), who originally proposed this technique, have shown convincingly that consistent discrimination peaks in the raw data are preserved by this data transformation procedure.

The labelling data obtained in this experiment are entirely consistent with the absence of discrimination peaks, in that the labelling curve represents a very smooth transition from 'high' judgements in early rises to 'low' judgements in late rises. The sudden cross-over which can be taken as indicative of underlying tonal categories does not materialize.

The second important characteristic of the discrimination data that has to be pointed out is that average item discrimination is extremely low, i.e. 36%. This indicates that informants generally do not discriminate between the stimuli in the discrimination items. This conclusion is reinforced by the results of a response coincidence analysis, which indicate that two groups of subjects have to be distinguished. The majority of subjects (60%) are classified as non-discriminators in that their average item discrimination is only 22%. The remaining subjects belong to the discriminator group, in which average item discrimination amounts to 49,5%. Taken together, these data clearly suggest that informants do not reliably discriminate between the stimuli taken from the alignment continuum.

In chapter 4, these observations were interpreted as a strong indication of the fact that the rise-alignment continuum is associated with a single alignment category. Consequently, it was argued that the precise location of the rise in the hat pattern in not essential to the perceptual and linguistic characterization of this pattern. In chapter 6, it was found that the just noticeable difference of rise-alignment was considerably larger than expected, i.e. 110 msec. As a result, the possibility has to be considered that the absence of discrimination peaks may have been caused by the fact that the gap sizes in the experimental design were below the discrimination threshold. Consequently, any discrimination peaks cannot materialize, since the differences in alignment are psycho-acoustically too small to be perceived.

In order to investigate this point further, a separate discrimination experiment was carried out, in which the gap size was increased to a value near the discrimination threshold. The results of this experiment show no significant discrimination peaks in the discrimination curve. The general level of item discrimination increases quite substantially, but there is no indication of a category boundary. Thus, the results of this experiment confirm and reinforce the conclusions in chapter 4 that the rise-alignment continuum in the Dutch hat pattern is related to a single perceptual category. This result is entirely in agreement with results in Boves et al. (1984). In this experiment, informants were required to imitate clear instances of the hat and cap patterns in which the alignment of the rises differed. It was found that this alignment difference is not reflected in the imitations.

Fall-alignment discrimination in the hat pattern was investigated by means of a 2IAX discrimination task. The data for this experiment show a clear and significant discrimination peak at item number 11. The pitch movements in this item are located in such a way that the fall of the first stimulus is situated entirely within the syllable's vowel nucleus, whereas the endpoint of the fall in the second stimulus is located in the postvocalic consonant. This discrimination peak can be accounted for in terms of a category boundary in the continuum, which coincides with the segment boundary between the syllable nucleus and the postvocalic consonant. This indicates that one category consists of falls of which both turning points are entirely situated within the vowel nucleus, whereas the falls in the second category have their endpoint in the postvocalic consonant. Since the disctinction between movements A and C can only materialize at the end of utterance phonation, an additional contrast between a complete falling movement and a partial movement is implied. This is entirely consistent with a classification of falls in the GDI into a late and very late category which corresponds to a linguistic distinction between prominence-lending and non-prominence lending falls.

As far as fall-discrimination is concerned, it was also investigated whether it is necessary to postulate a hierarchical relationship between the location of the rise in the hat pattern and the categorization of falls. This was done by manipulating the location of the prenuclear rise in the stimuli. It was found that rise location has a small but significant influence on fall discrimination.

On the basis of the conclusions regarding the discrimination of rise- and fall-alignment, we would like to argue for a more general characterization of both hat and cap patterns in Dutch, in that a specification of the precise location of rising pitch movements is not essential to the identity of either pattern. Any systematicities in the location of rising pitch movements in these patterns can be accounted for by means of a lower level phonetic component that relates to the interactions between intonation and other prosodic features such as speaking rate, the presence of prosodic boundaries in the pattern and the distance between the accents (Silverman & Pierrehumbert, in press). In order to specify these interaction between intonation and the other components, a detailed analysis of the IPO stylizations in terms of these factors may provide fundamental insights.

The precise location of the falls in both patterns, however, is to the largest extent determined by the pattern involved. In the hat pattern, the endpoint of the fall has to be located within the syllable nucleus of the accented syllable. In the cap pattern, this endpoint is situated in the postvocalic consonant. Within these restrictions, which are essential to the perceptual identity of both patterns, other prosodic factors determine the fine phonetic detail.

Unlike the GDI, the model of intonation that is proposed here explicitly assumes that there is no one-to-one relationship between the acoustic detail of F0 curves and abstract intonation patterns of the language by recognizing the importance of the interaction between intonation and other factors such a speech rate, prosodic environment and segment identity. It is clear that quite a substantial amount of work has to be done in order to understand the precise mode of interaction between these different components. This comprehensive approach to intonation analysis is however likely to provide a more comprehensive understanding of the structure of intonation.

POSTSCRIPT

Experimental Work Associated with the Thesis.

These postscript sections consist of short reports on the experimental work that has been carried out within the framework of this thesis. All the experiments that are described are in some way or another related to the main experiments in the body of the text, either as pilot experiments or as further explorations. Because they are not essential to the main topic of this thesis, they are presented separately.

CHAPTER 8

Categorical Perception and the Reality of Targets in Intonation.

- Two Pilot Experiments -

0. Introduction.

In this paper, the results of two perception experiments on Dutch intonation are reported. The purpose of these pilot experiments is twofold. On the one hand, we attempt to show that rises with a different location with respect to an accented syllable are perceived as categorically different by speakers. It is hypothesized that these categorical differences relate to the kind of phonological tonal segment that is associated with the syllable. On the other hand, preliminary evidence as to the phonetic nature of these segments is collected. The data presented here suggest that tonal segments should be phonetically characterized as targets (levels) rather than pitch movements (configurations).

It is generally accepted in tone sequence theories of intonation that intonation contours can be described in terms of a relatively small number of discrete units or segments. There is however no general agreement about the phonetic nature of these units. Phonological analyses postulate the existence of abstract tonal segments which account for the underlying linguistic structure of intonation patterns. From a phonetic point of view, these segments are regarded as high and low pitch levels. The more phonetically inspired analyses postulate pitch movements as units in the melodic description of pitch contours. Both points of view have been illustrated in chapter 2 of this dissertation with specific reference to the analysis of Dutch intonation.

1. Experiment 1: Categories in the Perception of Intonation.

It is one of the basic assumptions of a tone sequence theory that there are units or segments in terms of which intonation contours can be described. From this, two basic research questions can be derived. A first question relates to the reality of these segments. The second relates to the phonetic nature of these units. It is the former that will be dealt with first in this experiment.

1.1. Hypothesis.

Both the pitch movement and the target model differentiate between units, which differ from each other in the phonetic dimension of their alignment with respect to the syllable. The grammar of Dutch intonation distinguishes between an early rise 1, a late rise 3 and a very late rise 2. Gussenhoven's model postulates a difference between a HL and a LH tone, which in certain phonetic environments correlates with a phonetic difference between an early and a late rise. Furthermore, Ladd (1983a) postulates a feature [peak delay], which suggests that alignment differences are linguistically significant. Therefore we propose the following hypothesis:

> An early rise in an accented syllable is perceived categorically different from a late rise, their acoustic specifications otherwise being identical.

As to the location of the category boundary, we assume that it occurs between peaks situated 30 msec after vowel onset, and those situated at 90 msec. These values are the suggested standards for rise 1 and rise 3 as found in Collier (1970).

1.2. Experimental Design.

In order to test above-mentioned hypothesis, we have carried out two categorical perception experiments in which the location of an F0-rise on an utterance-initial syllable was the independent variable.

1.2.1. Stimuli.

Eight Dutch utterances were prepared by means of a slightly adapted version of the Holmes synthesis-by-rule system (Holmes et al, 1964), in which the vowel tables were altered for the acoustic specifications of Dutch. Furthermore, a number of Dutch consonants were modelled acoustically.

1.2.1.1. Segmental Structure.

All sentences had a similar syntactic/semantic structure in that they consisted of an Argument/Predicate/Argument-combination as in (1)

(1) Joop neemt lessen Russisch

A1	Joop
P	neemt
A2	lessen Russisch

Furthermore, A1 had a similar segmental structure in all utterances. It consisted of an initial voiced consonant, followed by a [-high] vowel and a voiceless postvocalic consonant. The vowel was chosen to be [-high] in order to keep the influence of intrinsic vowel pitch constant (Lehiste, 1970). Finally, the durational characteristics of the experimental syllable (A1) in the stimuli were kept constant:

(2)	Consonant	Vowel	Consonant
	[+voice]	[-high]	[-voice]
	90	140	70 msec

1.2.1.2. Intonation Structure.

All the utterances were provided with an identical intonation contour, which was specified with the values in figure 8.1. Formally, the pitch contour was the realization of a flat hat pattern:

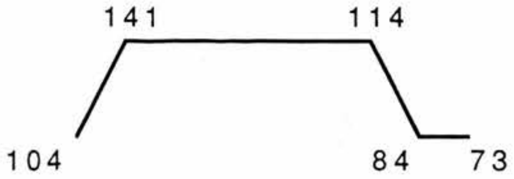


Figure 8.1: Acoustic specification of the pitch contours in the experimental utterances. Values are given in Hz and declination is omitted.

From these utterances with canonical intonation contour, the experimental stimuli were derived by shifting the position of the initial rise to the right in steps of 10 msec, leaving the remaining part of the contour unchanged. Hence, eight pitch movement alignment continua were obtained with each 21 different rise locations. These are schematized in figure 8.2:

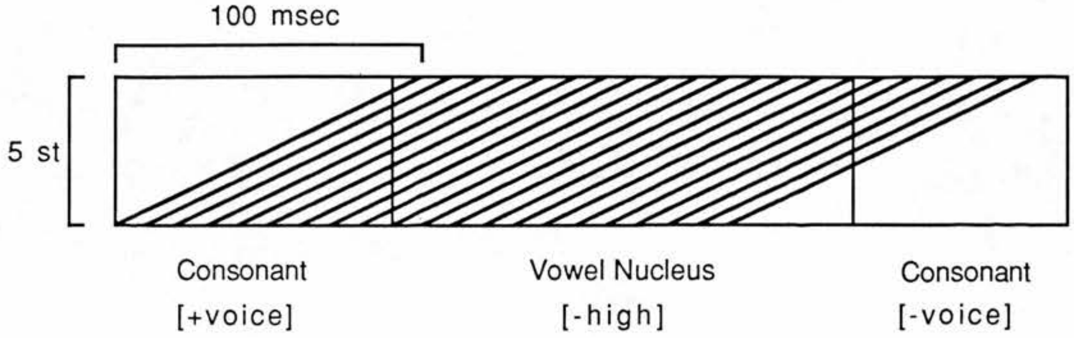


Figure 8.2: Acoustic specification and alignment of pitch rises on the experimental syllable. Rises are shifted to the right in steps of 10 msec.

Thus, a total number of 168 stimuli were obtained, i.e. 8 sentences with each 21 rise locations.

1.2.2. The Experimental Paradigm.

The stimuli were used in a 3-step and 4-step AX discrimination task. Informants were presented with two stimuli from the same continuum and were asked to judge whether the stimuli were identical or not.

For the purpose of this pilot experiment, an AX algorithm was preferred, since it does not impose heavy demands on the informants' memory (Repp, 1984). Furthermore, the AX-algorithm was slightly adapted. Ideally, for each items in which the stimuli are physically different in their acoustic characteristics, the test should contain an item in which the stimuli are acoustically identical. Judgements on these identical items can then be used to calculate the informants' response bias (Wood, 1976). This design principle was omitted since it reduces the number of items in the test by a factor of two. Furthermore, Ainsworth & Lindsey (1986) have shown that it is possible to obtain a discrimination function, even without the present of these reference items.

The stimuli were paired into AX-discrimination items with two different gap sizes. In the 3-step task, there was a 30 msec gap between the locations of the rises in both stimuli. In the 4-step task, the physical difference was 40 msec.

1.2.3. The Test Tapes.

The items in each task were randomized and recorded on tape. The interstimulus interval was 500 msec. The interval between the items was 3000 msec. Each item was preceded by an identifier in order to orientate the subjects in the test and to make comparison between succeeding items difficult.

In addition, 5 fillers were added at the beginning and the end of each tape. The judgements for these items were not taken into account in the calculation of the final results. Total duration of each tape was approximately 20 minutes.

1.2.4. Informants.

28 native speakers of Dutch were recruited in the Department of Germanic Philology of Antwerp University. 18 students took part in the 4-step and 10 in the 3-step discrimination task. All subjects took part on a voluntary basis and no informant took part in both tasks.

1.2.5. Procedure.

Informants were seated in a quiet language laboratory in which individual adjustment of the volume level was possible. They were told that they were going to take part in a perception experiment on Dutch intonation. Subjects were asked to concentrate on the initial rise of the contour in both stimuli of the items and they were asked to indicate whether these rises were identical or not. To keep record of their judgements, they were provided with a scoring sheet on which they could indicate the judgement of their choice for each item in terms of the labels 'same' and 'different'.

Before the start of the experiment, subjects listened to a demonstration tape with several repetitions of all utterances. This was done to familiarize them with the computer voice. Subsequently, the test tape was run without interruption.

1.3. Results.

For the 3-step discrimination task, a total of 1440 observations were obtained. For the 4-step task, we recorded 2448 judgements. These results are summarized in figure 8.3:

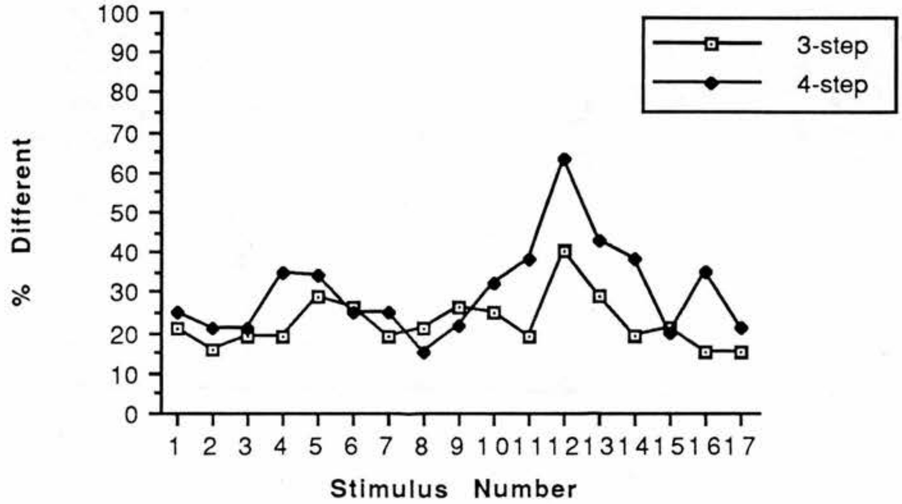


Figure 8.3: Percent different judgements (y-axis) as a function of item number (x-axis) in the 3-step and 4-step discrimination task. Each data-point in the 3-step task represents 80 observations. Those from the 4-step task represent 144 judgements.

Visual inspection of the judgement distribution in both discrimination tasks indicates that the majority of informants judge the stimuli in the discrimination items to be equal. It can however be noted that there is a considerable peak in the percentage of different judgements at item 12. For the 3-step task, 40% of the informants judge the stimuli of this item to be different (the average for the other items is 16%). This discrimination peak is even more outspoken in the 4-step task, in which 64% of the subjects judge the stimuli to be different (average on the other items is 28%). These results suggest that discrimination between the stimuli greatly improves at item 12 as compared to the other items in the task.

In order to investigate whether this deviation of judgements on item 12 is statistically relevant, we computed a χ^2 in which the average judgements on all the items were treated as the expected values, and the judgements on item 12 as the observed values. The results indicate that the observed judgements on item 12 do not deviate very significantly from the expected judgements in the 3-step task ($\chi^2 = 5.76$, d.f. = 1, $p < 0.025$). In the 4-step task however, judgements on item 12 deviate from the average at a highly significant level ($\chi^2 = 36.30$, d.f. = 1, $p<0.001$).

Taking into account these observations, the following picture of informant responses appears. Subjects judge the stimuli in all items but one to be equal. In the 4-step discrimination task, a statistically highly significant improvement of discrimination responses can be observed in item 12. This indicates that the rises in the stimuli of item 12 belong to categorically different classes. In other words: the discrimination peak can be explained by the presence of a category boundary in the rise alignment continuum. This boundary is located between the rises in the stimuli of item 12, which is illustrated in figure 8.4:

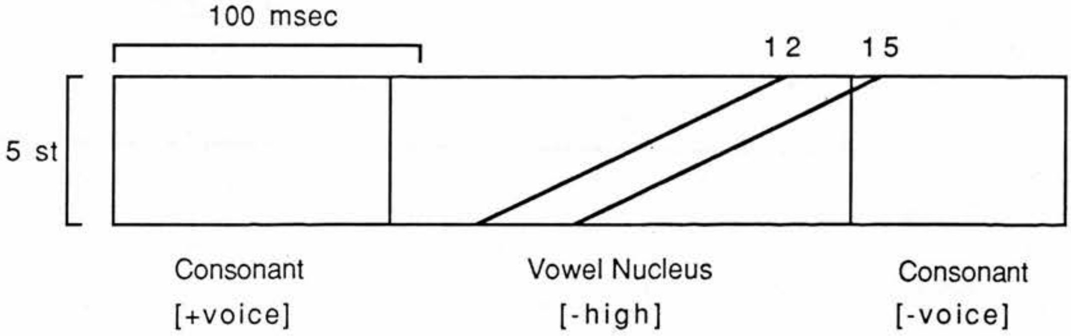


Figure 8.4: Schematic representation of the category boundary between the rise locations of item 12 and 15.

In all other items, the rises of the stimuli were taken from the same class, and consequently, informants could not tell them apart.

It should be observed that the categorization found here does not correspond to the IPO distinction between early, late and very late rises. Our first category (locations 1-12) contains rises which the IPO-model would subdivide into two melodically different categories: i.e. rise 1 and 3. The boundary in these data corresponds rather to the IPO distinction between rise 3 and rise 2.

Instead of a 3-way distinction in pitch movement alignment, the perceptual data presented here suggest that a two-way distinction is more appropriate. This distinction can be accounted for in terms of a phonological difference in the associated tonal elements. In the case of an early rise, we hypothesize that the rising pitch movement corresponds to an underlying HIGH tonal segment. In these rises, it is the F0 endpoint that is marked. In the late rises however, it is a LOW tone that is associated with the accented syllable. Hence, late rises mark the F0 onset. This interpretation is supported by the acoustic characteristics of both types of rises, in that it is always the hypothesized F0-target which is located in the vowel nucleus of the syllable, whereas the non-targets are situated in the pre or postvocalic consonants and as such are not perceivable if the consonant is for instance voiceless.

1.4. Conclusion.

The results of this experiment indicate that F0 rises are perceived as categorically different by informants, depending on their location with respect to the nucleus of the accented syllable. Instead of a three-way distinction, we suggest a two-way categorization in terms of early and very late rises. It is furthermore argued that it is not the location of the whole pitch chunk which is relevant to this categorization, but rather the location of the phonologically marked F0-targets.

2. Experiment 2: Targets vs. Pitch Movements.

The evidence presented in experiment 1 suggests that there are categorical differences in the perception of F0 rises. The experimental set-up did however not allow us to draw conclusions about the phonetic natures of the tonal segments which were hypothesized to correlate with this categorization. therefore, a second experiment was carried out, in which the F0 onset and endpoint of pitch movements were varied for the category of early rises, in which the H-target was assumed to be phonologically relevant. The results of this experiment indicate that tonal segments should be regarded as targets rather than pitch configurations.

2.1. Hypothesis.

The predictions of the pitch movements and target models concerning the perceptual consequences of variations in onset and offset of F0 rises are essentially different.
The IPO intonation model predicts that the perceptual impressions of two rises are different if their excursion size is different, provided that the variation of excursions exceeds a perceptual threshold of ca. 3 semi-tones ('t Hart, 1981). This observation was used as evidence that it is the pitch movement as a whole which is to be regarded as the basic perceptual unit in intonation. It follows from this interpretation that it should be irrelevant which end of the pitch movement is varied. In other words: if the endpoint of a rise is raised by more than 3 semi-tones, the two pitch movements can be expected to be perceptually different. Conversely, if the

onset of a rise is lowered by more than 3 semi-tones, the resulting perceptual impressions are not equivalent either.

In order for a target approach to be consistent, it will have to predict that the sensitivity of informants to variations in F0 onsets and offsets crucially depends on the phonological nature of the underlying tonal segment and on which part of the rise is manipulated. In other words, if an accented syllable is associated with a H tone, small variations in the F0 of this target should be expected to be clearly perceivable, whereas variations in the F0 onset of the movement towards this target should not really have any perceptual consequences. This is because it is the H-target which is linguistically significant and not the pitch movement as a whole. It is precisely this prediction of the target approach which we will be evaluated in this experiment.

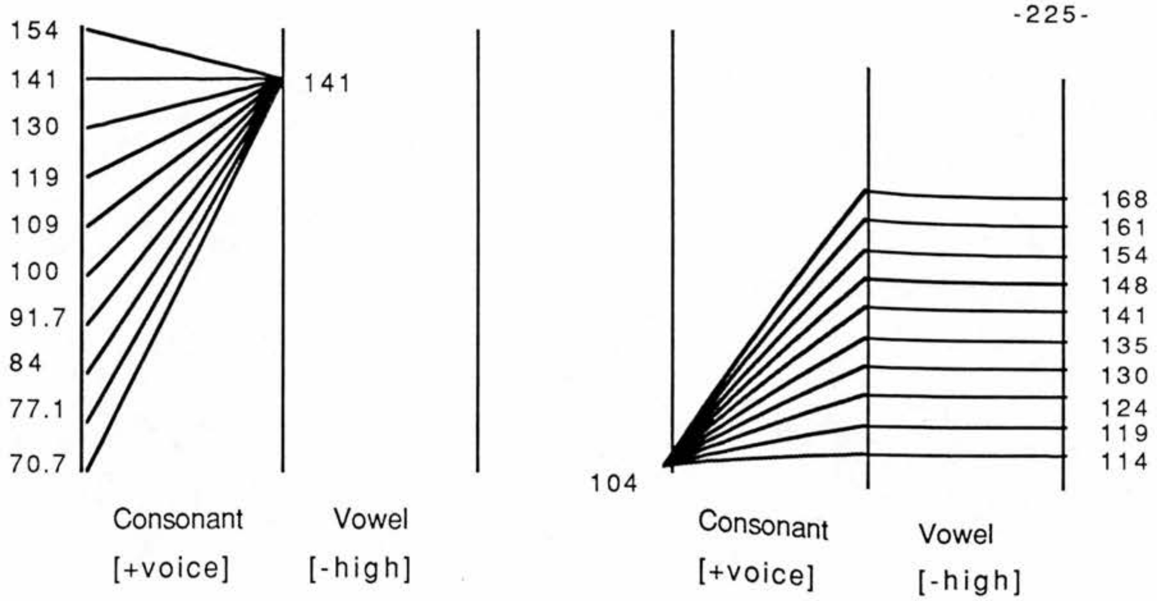
2.2. Experimental Design.

To test the hypothesis, a discrimination experiment was carried out, in which the onset frequency and endpoint frequency of early rises were manipulated on utterance-initial syllables. The rises were embedded in a flat hat pattern.

2.2.1. Stimuli.

4 Dutch utterances were synthesized by means of the Holmes synthesis-by-rule system. The segmental characteristics were identical to the utterances described in experiment 1. So was the specification of the intonation contour.

From these utterances, the experimental stimuli were derived by varying the frequencies of the F0 onset and F0 endpoint, as schematized in figure 8.5:



Condition 1A

Condition 1B

Figure 8.5: Schematic representation and acoustic characteristics of the experimental pitch movements in condition 1A and 1B. In condition 1A, the interval between the F0 onsets is ca 1.5 semi-tones. In condition 1B, the interval between F0 endpoints is ca. 0.75 semi-tones. All figures are given in Hz.

In condition 1A, a fixed F0 endpoint of 141 Hz was implemented at 10 msec after vowel onset. The onset of the movement towards this H-target was varied in 10 steps of ca. 2.60 semi-tones. Hence, 40 stimuli were obtained (10 x 4 sentences).
In condition !b, the F0 onset was fixed at 104 Hz, situated at the first frame of the initial voiced consonant. The F0 endpoint of the rise was varied in steps of 1.33 semi-tones, yielding a total of 40 stimuli (10 x 4 sentences).

2.2.2. Experimental Paradigm.

As in experiment 1, an AX discrimination algorithm was used: the stimuli from each phonetic continuum were paired into AX items, the difference between the values of the independent variable being 3 steps. This is to say that in condition 1A, there was a 4.5 semi-tone difference between the onset frequencies of the movements in stimuli A and X. In condition 1B, the difference between the endpoint frequencies of the rises in A and X amounted to 2.5 semi-tones.

2.2.3. The Test Tape.

All the items were randomized 4 times and recorded on tape. The interstimulus interval was 500 msec, whereas the items were separated by a 3000 msec pause. Each item was preceded by an identifier to orientate the subjects in the task.

In addition, 5 fillers were added at the beginning and end of the test, the judgements on which were ignored in the calculation of the results. Total duration of the tape was ca. 20 minutes.

2.2.4. Informants.

20 native speakers of Dutch of the Department of Germanic Philology of Antwerp University took part on a voluntary basis. Neither of these informants had taken part in the discrimination tasks of the previous experiment.

2.2.5. Procedure.

The subjects were told that they were going to take part in a perception experiment on Dutch intonation. The were asked to concentrate on the initial rise of the stimuli in each item and were required to indicate whether they could hear a difference between them or not. The listening conditions were identical to those in experiment 1 and the subjects were also given a short practice session in order to get accustomed to the computer voice.

2.3. Results.

In this experiment, a total of 1600 observations were obtained, i.e. 800 for each condition. These results are summarized in figure 8.6:

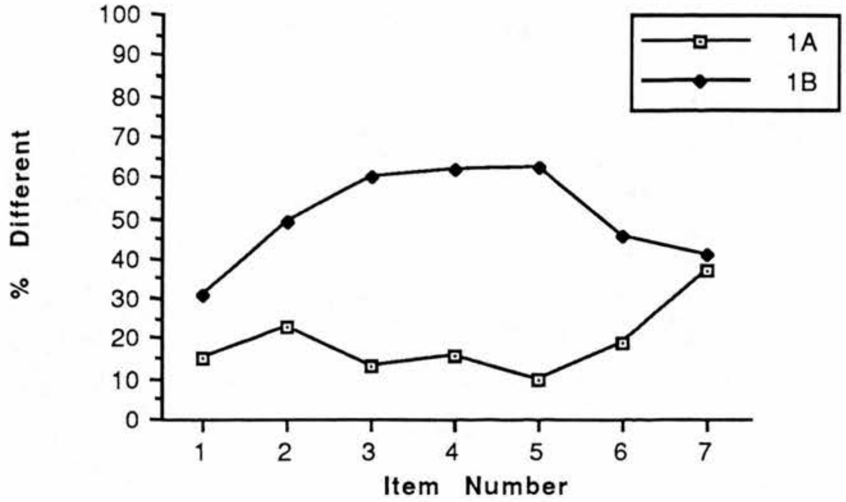


Figure 8.6: Percent different judgements (y-axis) as a function of item number (x-axis) in the two experimental conditions.

2.4. Discussion.

From figure 8.6, it appears that the distribution of judgements in conditions 1A and 1B are different. Calculation of χ^2 suggests that this difference in judgements is statistically significant ($\chi^2 = 8.53$, d.f. = 1, $p < 0.005$).

2.4.1. Condition 1A.

In condition 1A, the values of F0 onset of the pitch movement towards a fixed H-target were varied. These variations were not perceived by informants: on average, 81% of the subjects judged both stimuli in the items to be equal. This insensitivity to onset differences is so outspoken that informants did not hear a differences between a falling movement towards the target and a rising movement (item 1). For this item, 84% of equal judgements were obtained. In addition, we should like to emphasize that the actual physical difference between the onset value in all items was 4.50 semi-tones, which can be regarded as well above the perceptual threshold of 3 semi-tones ('t Hart, 1981). This observation strongly suggests that it is not the pitch movement as a whole which is perceptually relevant, but rather the target value only.

The reader should note the considerable rise in the different judgements at item 7. This peak is probably related to the influence of the very low F0 onset, which affected the segmental quality of the initial consonant: it sounded as if it was produced with creaky voice, and this may have influenced informants judgements.

2.4.2. Condition 1B.

The interpretation of the results in this condition is slightly more complex. Our hypothesis predicted that informants would judge the stimuli in ALL items to be different. This is clearly not the case. Figure 8.6 indicates an unequal distribution of the judgements. Two groups seem to materialize: on the one hand, there are items 1, 2, 6 and 7 in which 41% of the informants judge the stimuli to be different. In items 3, 4 and 5 we observed 62% different judgements. This indicates that discrimination is determined by the relative positions of the endpoints in the continuum, or in other words: F0 endpoints are perceived in terms of categories. If informants have to compare stimuli which belong to the same category, they tend not to hear the difference between them. If the stimuli are taken from different categories, discrimination improves considerably. As to the statistical significance of increased discrimination in items 3, 4 and 5, χ^2 indicates a moderate significance ($\chi^2 = 5.62$, d.f. = 1, $p < 0.010$).

As to the reference point with which peak height seems to be judged in this type of contour, the pitch level of the following syllable nucleus seems to be an important factor. In our stimuli, the continuation level was 135 Hz and stimuli 3,4 and 5 had peak values which belonged to different categories with respect to this 135 Hz-level, as summarized in table 8.1:

	Stimulus A	Stimulus X
3	154	130
4	148	124
5	141	119

Table 8.1: Peak height (in Hz) in stimuli A and X in items 3, 4 and 5. Continuation level in all stimuli was 135 Hz.

In the other items, peak values in both stimuli were either above or below this continuation level.

Taking all these observations into account, we can say that informants' perceptual responses to variations in F0 onset and endpoint is fundamentally different. Variations in F0 onsets are not perceived, although differences in onset values were above the threshold level. Variations in F0 endpoints had a non-uniform effect on judgements: peak height is evaluated with respect to the continuation level of the contour, and informants do perceive differences if the peaks belong to different categories, even though these differences in peak values were below the threshold level.

This differential response of informants can be accounted for by the fact that the H-target is linguistically important, and not the pitch movement as a whole. This interpretation is also supported by preliminary examination of very late rises, in which a low target is assumed to be marked. Similar variations of F0 onsets and endpoints also lead to a differential response: if the onset (low target) was manipulated, a vast majority of informants judged the rises to be different. Variations in the F0 endpoint caused subjects to judge the rises to be identical. This is in fact the reverse response behaviour to that presented in this experiment, but it is consistent with the interpretation that subjects are more sensitive to F0 variations related to the linguistic target than to other F0 variations.

3. Conclusion.

In these experiments, we have shown that informants' discrimination of F0 reflects categorical organization in terms of the location of the F0 target which is associated with the rise. For early rises, it is the H-target which is relevant. For late rises, the L-target is important. This interpretation is confirmed by the fact that in early rises, variations in the F0 onset are not relevant to perception, even if it involves a change in the identity of the movement (rise vs. fall), and even if these variations are above the perceptual threshold. Furthermore, late rises, which are assumed to mark a L-target, elicit similar differential discrimination responses. This indicates that a characterization of pitch movements in terms of targets does not only enable phonological generalizations to be expressed, but also accounts for the observed perceptual responses of informants.

CHAPTER 9

Categorical Perception of Vowel Duration in German. -An Investigation of Phonetic Prototypes-

0. Introduction.

In categorical perception experiments on speech as well as non-speech phenomena, it has often been observed that the presentation order of the stimuli in the discrimination items has a significant influence on subjects' discrimination performance (Batliner & Schiefer, 1987). This effect is generally known as the 'order effect'. In chapter 4, a highly systematic order effect was observed, which can be appropriately characterized as 'order effect reversal'. It was hypothesized that this reversal can be accounted for in terms of phonetic prototypes, if it is assumed that phonetic categories can be realized as ideal, good and poor examples of the category.
In this perspective, it can be predicted that if the two stimuli in a discrimination item are ideal realizations of the phonetic prototype, the order of presentation of these stimuli has no significant effect on discrimination performance. If the two stimuli belong to the same phonetic category, but are not realizations of the prototype, the order effect shows up. It is predicted that the precise nature of the order effect is related to which stimulus is presented first. If the presentation order involves a shift away from the prototype, i.e. the better example of the category is presented first, discrimination is expected to be high. If on the other hand the presentation order of the stimuli involves a shift towards the phonetic prototype, i.e. the poorer example of the category is presented first, discrimination is expected to be lower. These predictions are not only consistent with the data for rise-alignment in chapter 4 and the perception data on intonation reported in Kohler (1987) and Batliner & Schiefer (1987), but also account for the observations on the perception of segmental aspects of speech in which the presentation order of the stimuli was controlled (Repp et al., 1979, Uselding, 1977).

1. Experimental design.

In order to test the predictions formulated above, a 2IAX discrimination experiment was carried out in which subjects were required to discriminate between pairs of stimuli that were taken from three continua of vowel duration in German. On a separate occasion, the subjects participated in a labelling experiment, in which they were asked to identify the stimuli in terms of one of three word-pairs.

1.1. Construction of the stimuli.

The stimuli for this experiment were derived from three German words with a long vowel duration: Staat, Strafen and Staatlich. These words were spoken 10 times by a male native speaker of German, who realized them with a long [a] vowel. In his deliveries, the informant made an effort to make the vowel as long as possible while without becoming unnatural. From these deliveries, one realization of each word was chosen which was natural and fluent in all respects.

The realizations were digitized at Fs = 20,000 Hz at Fc = 8,000 Hz and segmented. The result of this segmentation is given in table 9.1:

Staat	ʃtɑt	ʃt	175
		ɑ	
		t	100
Staatlich	ʃtɑtliç	ʃt	175
		ɑ	
		tliç	435
Strafen	ʃtʀɑfn	ʃtʀ	240
		ɑ	
		fn	195

Table 9.1: Durational characteristics of the three experimental words.

After obtaining information about the segment clusters in the words, the a-vowel of 'Staat' was chosen as the experimental vowel: it had a duration of 280 msec in the original recording. From this vowel, 26 versions were derived by shortening the original in steps of 10 msec starting from the right. All the segment clusters and vowels were stored in a dictionary, which was used to synthesize the stimuli for the experiment by means of a waveform recombination technique (Dryden & Mcleod, 1988). In this technique, the cluster [ʃt] from the origninal recording was recombined with the 27 vowels and the resulting waveforms were combined with the original [t]. Consequently, a continuum of vowel length between long and short [a] was obtained, in which the long version sounded like 'Staat', whereas the short version sounded like 'Stadt'. By applying the same technique to 'Staatlich' and 'Strafen', two additional continua were obtained, i.e. 'Staatlich' vs 'Stattlich' and 'Strafen' vs. 'Straffen' respectively. This procedure yielded 81 stimuli (i.e. 27 x 3 continua).

The advantage of this synthesis technique is that very natural stimuli of high quality can be obtained in which all the segmental characteristics are identical, except for the vowel duation. Moreover, the vowel quality in the three continua was identical, since the three vowel continua were derived from the same vowel.

1.2. Item Construction.

For the discrimination task, the stimuli in each continuum were combined into variable standard 2IAX discrimination items with a 30 msec gap size between the stimuli. In addition, a set of control items were used, in which the durational characteristics of the vowels in both stimuli were identical. In all items, the stimuli were presented in both long-short (AB) and short-long (BA) presentation order. Since every item occurred only once in the test, 225 discrimination items were obtained.

1.3. Test Tapes.

The items of the discrimination task were recorded on tape in different presentation blocks, according to the wordpair involved. In each presentation block, the AB, BA and AA items occurred in random order. There were no identifiers between the items, but after presentation of 10 items, the informants heard an orientation signal. The inter-stimulus interval was 500 msec, whereas the inter-item interval amounted to 2500 msec. For the discrimination task, the stimuli representing each wordpair were randomized and recorded on tape with an interstimulus interval of 4000 msec. After every set of 10 stimuli, informants heard an orientation signal.

1.4. Informants.

48 native speakers of German attending the 'Zomercursus Nederlandse Taal- en Cultuur' at the 'Economische Hogeschool Limburg' took part in the experiment. They participated on a voluntary basis.

1.4. Procedure.

The informants were seated in a quiet language laboratory, in which the volume level of the headphones was individually adjustable. They were told that they were going to take part in a perception experiment investigating vowel duration in German. In the discrimination experiment, they had to compare the stimuli in the items and were asked to judge whether they were the same or different. In the identification experiment, the subjects were presented with the individual stimuli from each continuum and were asked to give two judgements. In the first instance, they were required to identify the stimuli as either 'Stadt'/'Staat', 'Stattlich'/'Staatlich' or 'Straffen'/'Strafen'. Secondly, the informants had to indicate whether the vowel duration in each stimulus was 'typical' or 'atypical' for the word they recognized.

The informants were given the following instructions :

> Das Experiment, an dem Sie teilnehmen werden, handelt um die Wahrnehmung langer und kurzer Vokale in deutschen Wortpaaren wie z.B. 'Stadt/Staat'. Sie hören eine Reihe von Paaren, an denen die Vokallänge künstlich manipuliert worden ist. Ihre Aufgabe ist es, für jedes Paar zu beurteilen, ob Sie einen Unterschied zwischen den beiden Teilen des Paares hören. Ihr Urteil geben Sie für jedes Wortpaar an, indem Sie ein Kreuz ins zutreffende Kästchen (G = gleich, V = verschieden) machen.
>
> Manche Wortpaare bestehen aus zwei Versionen von 'Staat', andere aus zwei Versionen von 'Stadt', wieder andere sind zusammengesetzt aus Beispielen beider Wörter. Auch wenn es um 2x 'Staat' bzw. 2x 'Stadt' geht, kann es trotzdem manchmal sein, daß Sie einen Unterschied feststellen. Hören Sie bitte genau zu! Bitte treffen Sie für ALLE Wortpaare eine Entscheidung, auch wenn Sie sich nicht ganz sicher sind.
>
> Die Wortpaare sind auf dem Tonband in Gruppen von 10 Paaren angeordnet, was der Ausmachung des Testbogens entspricht. Nach jedem Paar folgt eine Pause von etwa 3 Sekunden, währenddessen Sie Ihr Urteil abgeben sollen. Nach jedem 10. Paar hören Sie einen Doppelton sowie eine etwas längere Pause. Falls Sie innerhalb einer Gruppe von 10 Paaren etwas ausgelassen oder irgendwie falsch angekreuzt haben, so machen Sie bitte nach dem Doppelton mit der neuen Zehnergruppe weiter.
>
> Nach den Paaren 'Stadt/Staat' hören Sie ganz ähnliche Wortpaare 'stattlich/staatlich' und 'straffen/strafen'. Diese Paare sind ebenfalls in Gruppen von 10 Paaren auf dem Tonband angeordnet. Sie sollten genau wie bei den Paaren 'Stadt/Staat' die Entscheidung treffen, ob die beiden Teile des Paares gleich oder verschieden sind.
>
> Im zweiten Teil des Experiments hören Sie keine Wortpaare, sonder nur einzelne Beispiele der Wörter des ersten Teils. Diesmal haben Sie zwei Urteile abzugeben. Erstens geben Sie für jedes Wort an, ob es sich z.B. um 'Stadt' oder 'Staat' handelt. Zweitens sollen Sie gleichzeitig beurteilen, ob die Länge des Vokals typisch ist für das jeweilige Wort, oder ob der Vokal zu lang bzw. zu kurz ist. Diese Urteile geben Sie wie im ersten Teil an, indem Sie ein Kreuz ins zutreffende Kästchen machen.

BEISPIEL: Sie hören ein Wort, das deutlich als 'Stadt' einzustufen ist, obwohl der Vokal künstlich verlängert (d.h. Richtung 'Staat') worden scheint. Dieses Urteil geben Sie so an:

Stadt	Staat	zu kurs	typisch	zu lang
X	°	°	°	X

Bitte geben Sie für JEDES Wort ZWEI Urteile ab:

1. Wortidentität, 2. Vokallänge.

Vielen Dank für Ihre Mitarbeit!

2. Results and Discussion.

In the discrimination experiment a total of 10,800 observations were obtained, whereas the identification experiment yielded 7,776 judgements. The average discrimination curve is given in figure 9.1. The identification data are summarized in figure 9.2.

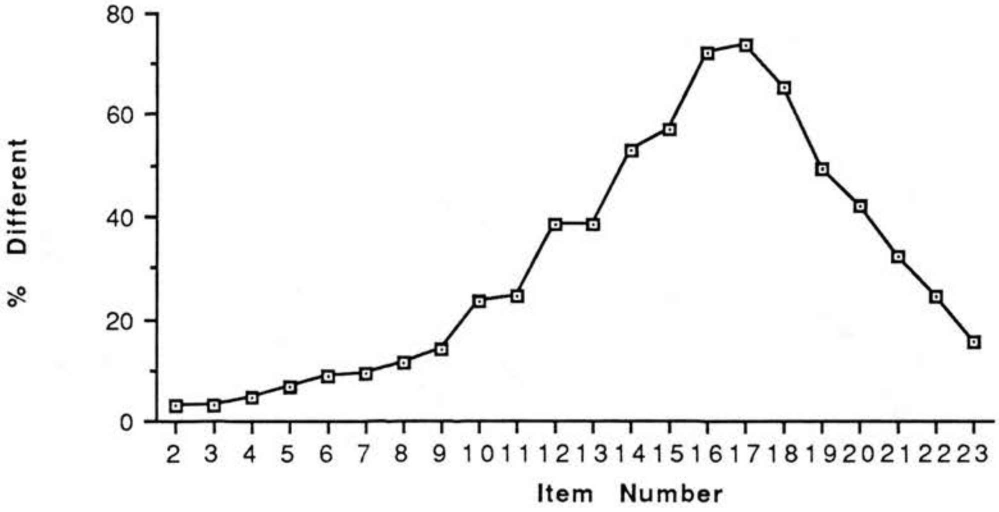


Figure 9.1 : Average discrimination curve for all the items in the experiment. The function is smoothed by 2 points. Total number of observations is 8,496.

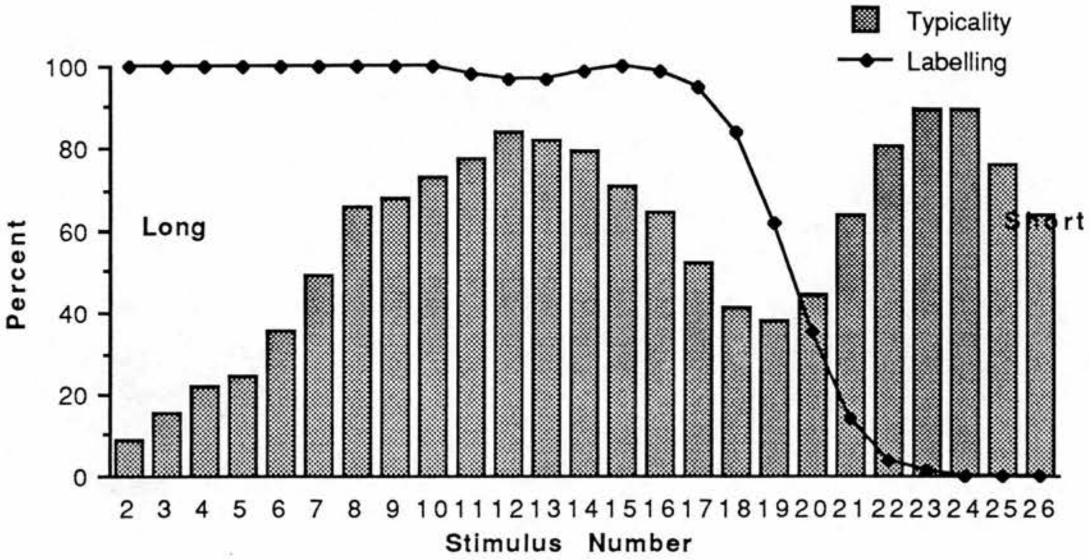


Figure 9.2 : Percentage of Identification and typicality judgements for the stimuli in the identification experiment as a function of decreasing vowel length. Both data sets are smoothed by 2 points. Total number of observations is 5,184.

It can be seen from figure 9.1 that there is a clear discrimination peak at item number 17. This indicates that an equal physical difference in vowel duration of 30 msec is not perceived uniformly throughout the continuum. At the extremes of the continuum, this difference is not discriminated, whereas it is clearly perceived around item number 17. The first stimulus of this item contains a vowel with a duration of 90 msec, whereas the vowel in the second stimulus has a duration of 120 msec. Informants discrimination behaviour in this experiment can be accounted for by assuming that the stimuli at item 17 represent different categories of vowel duration in German. This interpretation is supported by the labelling function in figure 9.2, which is characterized by a sudden cross-over between long and short vowels, which corresponds to the location of the discrimination peak.

Figure 9.2 also indicates that the the percentage of 'typical' judgements varies throughout the continuum. The vowels are regarded as most atypical at the extremes of the vowel length continuum and in the area where the cross-over in labelling judgements occurs. It suggests that the long vowels have an ideal value somewhere in the region of 170 msec (stimulus 12). The optimal duration of short vowels ranges between 50 and 60 msec (stimuli 23-24).

In order to obtain information about the predictions of the hypothesis that the order effect shows up in the discrimination items with atypical vowel durations only, the presentation order of the stimuli in the items was singled out as a separate variable, by quantifying OE as AB-BA. The result of this quantification is given in figure 9.3:

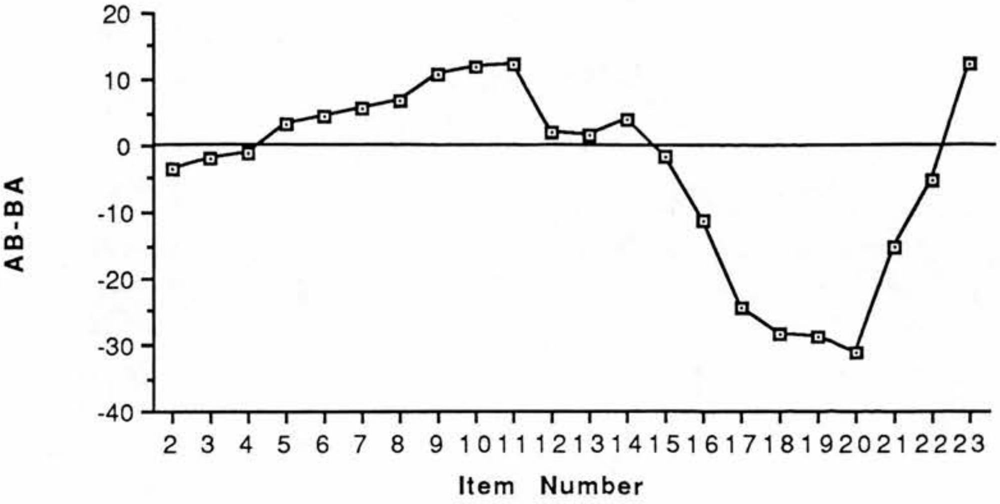


Figure 9.3: Quantification of the order effect in the experimental items as AB-BA.

Inspection of the quantified curve reveals substantial changes in the order effect. In items 2,3 and 4 the OE is negative and becomes positive in items 5-14. In items 15-22, the order effect is negative and it is positive in item 23. In order to evaluate the significance of this variation, a single sample t-test was carried out in which the quantifications of the order effect were compared to a mean of zero (i.e. no order effect). The statistic is not significant in that $p < 0.450$. It can be concluded from this that the predictions of the prototype hypothesis are not confirmed by the data presented here.

3. Conclusions.

The results of this experiment clearly show the existence of a categorical boundary between long and short vowels in German. Indicative of this category boundary is a clear discrimination peak in the discrimination curve, which coincides with a sudden cross-over in labelling judgements in the same area of the continuum. Furthermore, it was found that the ideal length of long vowels is 170 msec, whereas this for short vowels amounts to 50-60 msec. Although this finding is interesting in its own right it will not be further discussed here, since it is not directly relevant to the prototype hypothesis elaborated in chapter 4.

As far as the relationship between the presentation order of the stimuli in the items and discrimination is concerned, a substantial order effect is observed in figure 9.3. A statistical comparison between these quantifications and the values to be expected in the absence of an order effect reveals that the effect is not significant in this experiment. As a result, the prototype hypothesis is not substantiated by the results.

REFERENCES

Adriaens, L.M.H. (1984). *A Preliminary Description of German Intonation.*
IPO Annual Progress Report, vol. 19, pp. 36-41.

Ainsworth, W.A., Lindsay, D. (1986). *Perception of Pitch Movement on Tonic Syllables in British English.*
J. Acoust. Soc. Am., vol. 79, pp. 472-480.

Baker, W.J., Hogan, J.T., Rozsypal, A.J. (1988). *Response Coincidence Analysis: a Technique for Assessing Individual Difference in Response Styles.*
Journal of Phonetics, vol. 16, pp. 401-416.

Batliner, A., Schiefer, L. (1987). *Stimulus Category, Reaction Time, and Order Effect. An Experiment on Pitch Discrimination.*
Proc. Int. Con. Phon. Sci., Tallinn, vol.4, pp. 46-49.

Berg Van den, J.W. (1974). *Mechanism of the Larynx and the Laryngeal Vibrations.*
IN Malmberg (ed.), Manual of Phonetics. Amsterdam: North-Holland. pp. 278-308.

Bolinger, D.L. (1958). *A Theory of Pitch Accent in English.*
Word, vol. 14, pp. 109-149.

Borden, G.J., Harris, K.S. (1980). *Speech Science Primer. Physiology, Acoustics and Perception of Speech.*
Baltimore: Williams & Wilkins.

Boves, L., Ten Have, B.L., Vieregge, W.H. (1984). *Automatic Transcription of Intonation in Dutch.*
IN Gibbon, Richter (eds.), Intonation, Accent and Rhythm. Berlin/New York: de Gruyter. pp. 20-45.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective.*
Lund: Gleerup.

Catford, J.C. (1977). *Fundamental Problems in Phonetics.*
Edinburgh: University Press.

Clark, J., Yallop, C. (1990). *An Introduction to Phonetics and Phonology.*
Oxford: Basil Blackwell.

Cohen, A., 't Hart, J. (1967). *On the Anatomy of Intonation.*
Lingua, vol. 19, pp. 177-192.

Cohen, A., Collier, R., 't Hart, J. (1982). *Declination: Construct or Intrinsic Feature of Speech Pitch.*
Phonetica, vol. 39, pp. 254-273.

Collier, R. (1970). *The Optimum Position of Prominence Lending Rises.*
IPO Annual Progress Report, vol. 5, pp. 82-85.

Collier, R. (1972). *From Pitch to Intonation.*
Katholieke Universiteit Leuven, unpublished doctoral dissertation.

Collier, R. (1974). *Intonation from a Structural Linguistic Viewpoint: a Criticism.* Linguistics, vol. 129, pp. 5-28.

Collier, R. (1975a). *Perceptual and Linguistic Tolerance in Intonation.*
IRAL, vol. XIII, pp. 293-307.

Collier, R. (1975b). *Physiological Correlates of Intonation Patterns.*
J. Acoust. Soc. Am., vol. 58, pp. 249-255.

Collier, R. (1979). *Problems and Methods in the Study of Intonation.*
Annali della Scuola Normale Superiore di Pisa. Classe di Lettere e Filosofia. Serie III, vol. IX, pp. 357-365.

Collier, R. (1983). *Some Physiological and Perceptual Constraints on Tonal Systems.*
Linguistics, vol. 21, pp. 237-247.

Collier, R. (1987). *F0 Declination: the Control of its Setting, Resetting and Slope.*
IN Baer, Sasaki, Harris (eds.), Laryngeal Function in Phonation and Respiration.
Boston/Toronto/San Diego: College-Hill. pp. 403-421.

Collier, R. (1989). *Intonation Analysis: the Perception of Speech Melody in Relation to Acoustics and Production.*
IN Tubach, Mariani (eds.), Proc. European Conference on Speech Communication and Technology. Edinburgh: CEP.

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

Daniloff, R, Schuckers, G, Feth, L. (1980). *The Physiology of Speech and Hearing: An Introduction.*
Englewood Cliffs: Prentice-Hall.

Dryden, N. & Mcleod, I. (1988). *Speech Synthesis by Concatenation of Segmented Elements of Natural Speech.*
University of Edinburgh, Work in Progress, vol. 21, pp. 147-151.

Fant, G. (1960). *Acoustic Theory of Speech Production.*
The Hague: Mouton.

Flanagan, J.L., Saslow, M.G. (1958). *Pitch Discrimination for Synthetic Vowels.*
J. Acoust. Soc. Am., vol. 30, pp. 435-442.

Geel, R.C. van (1983). *Pitch Inflection in Electrolaryngeal Speech.*
Doctoral Dissertation, Univerity of Utrecht.

Gelfer, C.E., Harris, K. S., Baer, T. (1987). *Controlled Variables in Sentence Production.*
IN Baer, Sasaki, Harris (eds.), Laryngeal Function in Phonation and Respiration.
Boston/Toronto/San Diego: College-Hill. pp. 422-435.

Govaert, G.A., Van Katwijk, A. (1968). *Prominence as a Function of the Location of Pitch Movement.*
IPO Annual Progress Report, vol. 2, pp. 115-117.

Guilford, J.P. (1954). *Psychometric Methods.*
New York: McGraw-Hill.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents.*
Dordrecht: Foris Publications.

Gussenhoven, C. (1988a). *Adequacy in Intonation Analysis: The Case of Dutch.*
IN Van der Hulst, Smith (eds.), Autosegmental studies on Pitch Accent. Dordrecht: Foris. pp. 231-258.

Gussenhoven, C. (1988b). *Toonsegmenten in de intonatie van het Nederlands.*
GLOT, vol. 10, pp. 313-322.

Gussenhoven, C., Blom, J.G. (1978). *The Perception of Prominence by Dutch Listeners.*
Phonetica, vol. 35, pp. 216-230.

Hardcastle, B (1976). *Physiology of Speech Production: an Introduction for Speech Scientists.*
London: Academic Press.

Hart 't, J. (1976). *Psychoacoustic Backgrounds of Pitch Contour Stylisation.*
IPO Annual Progress Report, vol. 11, pp. 11-19.

Hart 't, J. (1979a). *Relations between Physical and Perceptual Aspects of Intonation.*
Annali Della Scuola Normale Superiore di Pisa, Classi di lettere e filosofia, Serie III, vol. IX, pp. 367-379.

Hart 't, J. (1979b). *Temporele kwantisering van spraak; een eerste exploratie.*
IPO rapport nr. 358.

Hart 't, J. (1981). *Differential Sensitivity to Pitch Distance, Particularly in Speech.*
J. Acoust. Soc. Am., vol 69, pp. 811-821.

Hart 't, J. (1984). *A Phonetic Approach to Intonation: from Pitch Contours to Intonation Patterns.*
IN Gibbon, Richter (eds.), Intonation, Accent and Rhythm. Berlin/New York: de Gruyter. pp. 193-202.

Hart, 't, J. (1986). *Declination Has Not Been Defeated - A Reply to Lieberman et al.*
J. Acoust. Soc. Am., vol. 80, pp. 1838-1840.

Hart 't, J., Collier, R. (1975). *Integrating Different Levels of Intonation Analysis.*
J. of Phonetics, vol. 3, pp. 235-255.

Hart 't, J., Collier, R. (1979). *On the Interaction of Accentuation and Intonation in Dutch.*
Proc. IXth Int. Congr. Phon. Sci., Copenhagen, vol. II, pp. 395-402.

Hart 't, J., Nooteboom, S.G., Vogten, L.L.M., Willems, L.F. (1981). *Manipulaties met spraakgeluid.*
Philips Technisch Tijdschrift, vol. 40, pp. 108-119.

Hatch, E., Farhady, H. (1982). *Research Design and Statistics for Applied Linguistics.*
Rowley/London/Tokyo: Newbury House.

Hess, W. (1982). *Algorithms and Devices for Pitch Determination of Speech Signals.*
Phonetica, vol. 39, pp. 219-240.

Hill, D.R., Reid, N.A. (1977). *An Experiment on the Perception of Intonational Features.*
Int. J. of Man-Machine Studies, vol. 9, pp. 337-349.

Hirano, M., Kakita, Y (1985). *Cover-Body Theory of Vocal Fold Vibration.*
IN Daniloff (ed.), Speech Science. London: Taylor & Harris, pp. 1-46.

Hollien, H. (1960a). *Some Laryngeal Correlates of Vocal Pitch.*
J. of Speech and Hearing Research, vol. 3, pp. 52-58.

Hollien, H. (1960b). *Vocal Pitch Variation Related to Changes in Vocal Fold Length.*
J. of Speech and Hearing Research, vol. 3, pp. 150-156.

Hollien, H., Moore, G.P. (1960). *Measurements of the Vocal Folds during Changes in Pitch.*
J. of Speech and Hearing Research, vol. 3, pp. 157-165.

Holmes, J., Mattingley, I., Shearme, J. (1964). *Speech Synthesis by Rule.*
Language and Speech, vol. 7, pp. 127-143.

House, D. (1985). *Implications of Rapid Spectral Changes on the Categorization of Tonal Patterns in Speech Perception.*
Lund University-Dep. of Linguistics Working Papers, vol. 28, pp. 69-89.

House, D. (1987). *Perception of Tonal Patterns in Speech: Implications for Models of Speech Perception.*
Proc. 11th Int. Congr. Phon. Sc., Tallinn, pp. 1-4.

House, D. (1990). *Tonal Perception in Speech.*
Lund: University Press.

ILS (1986). *Command Reference Guide V5.0.*
Signal Technology Inc.

Katwijk van, A. (1974). *Accentuation in Dutch. An Experimental Linguistic Study.*
Asse: Van Gorcum.

Klatt, D. H. (1973). *Discrimination of Fundamental Frequency Contours in Synthetic Speech: Implications for Models of Pitch Perception.*
J. Acoust. Soc. Am., vol. 53, pp. 8-16.

Kohler, K. (1987). *Categorical Pitch Perception.*
Proc. XIth Int. Con. Phon. Sci. Tallinn, vol. 5, pp. 331-333.

Ladd, D.R. (1983a). *Peak features and Overall Slope.*
IN Cutler, Ladd (eds.), Prosody: Models and Measurements. Berlin: Springer. pp. 39-52.

Ladd, D.R. (1983b). *Phonological Features of Intonational Peaks.*
Language, vol. 59, pp. 721-759.

Ladefoged, P. (1967). *Three Areas of Experimental Phonetics.*
London: Oxford University Press.

Laver, J. (1980). *The Phonetic Description of Voice Quality.*
Cambridge: University Press.

Laver, J. (in press). *General Phonetic Theory. An Introduction.*
Cambridge: University Press.

Lehiste, I. (1970). *Suprasegmentals.*
Cambridge, MA: MIT Press.

Leroy, L. (1984). *The Psychological Reality of Fundamental Frequency Declination.*
Antwerp Papers in Linguistics, nr. 40.

Lieberman, Ph. (1961). *Perturbations in Vocal Pitch.*
J. Acoust. Soc. Am., vol. 32, pp. 451-453.

Lieberman, Ph. (1968). *Intonation, Perception and Language.*
Cambridge MA: M.I.T. Press.

Macmillan, N.A. (1987). *Beyond the categorical/continuous distinction: A psychophysical approach to processing modes.*
IN: Harnad (ed.). Categorical Perception. Cambridge. pp. 53-87.

Macmillan, N.A., Kaplan, H.L. (1977). *The Psychophysics of Categorical Perception.*
Psychological Review, vol. 84, pp. 452-471.

Maeda, S. (1976). *A Characterization of American English Intonation.*
Unpublished Doctoral Dissertation, Mass. Inst. of Technology.

Maeda, S. (1979). *On the F0 Control Mechanisms of the Larynx.*
IN Boë, Descout, Guérin (eds.), Larynx & Parole. Institut de Phonétique de Grenoble, pp. 243-258.

Moore, B.C.J. (1973). *Frequency Difference Limens for Short Duration Tones.*
J. Acoust. Soc. Am., vol. 54, pp. 610-619.

Müller, J. (1837). *Handbuch der Physiologie des Menschen.*
Coblenz: Holscher.

Nabelek, I., Hirsch, I.J. (1969). *On the Discrimination of Frequency Transitions.*
J. Acoust. Soc. Am., vol. 45, pp. 1510-1519.

Nabelek, I.V., Nabelek, A.K., Hirsch, I.J. (1970). *Pitch of Tone Bursts of Changing Frequency.*
J. Acoust. Soc. Am. , vol. 48, pp. 536-553.

Nooteboom, S. G., Cohen, A. (1984). *Spreken en verstaan. Een nieuwe inleiding tot de experimentele fonetiek.*
Assen: Van Gorcum.

Odé, C. (1989). *Russian Intonation: A Perceptual Description.*
Amsterdam: Rodopi.

Pierrehumbert, J. (1979). *The Perception of Fundamental Frequency Declination.*
J. Acoust. Soc. Am., vol. 66, pp. 363-369.

Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation.*
Unpublished doctoral thesis: MIT.

Pierrehumbert, J., Steele, S. (1987). *How Many Rise-Fall-Rise Contours.*
Proc. XIth Int. Con. Phon. Sci. Tallinn, vol. 4, pp. 145-148.

Pierrehumbert, J., Steele, S. (1989). *Categories of Tonal Alignment in English.*
Phonetica, vol. 46, pp. 181-196.

Pijper de, J.R. (1984). *Modelling British English Intonation.*
Dordrecht: Foris Publications.

Pike, K.L. (1972). *General Characteristics of Intonation.*
IN Bolinger (ed.), Intonation. Harmondsworth, Penguin. pp. 53-86.

Pollack, I, Pisoni, D (1971). *On the Comparison between Identification and Discrimination Tests in Speech Perception.*
Psychonometric Sciences, vol. 24, pp. 299-300.

Repp, B.H. (1984). *Categorical Perception: Issues, Methods, Findings.*
In Lass (ed.), Speech and Language: Advances in basic research and practice, vol. 10. New York: Academic Press, pp. 224-322.

Repp, B.H., Healy, A.F., Crowder, R.G. (1979). *Categories and Context in the Perception of Isolated Steady-State Vowels.*
Journal of Experimental Psychology: Human Perception and Performance, vol. 5, pp. 129-145.

Rietveld, A.M.C., Gussenhoven, C. (1985). *On the Relation between Pitch Excursion Size and Prominence.*
J. of Phonetics, vol. 13, pp. 299-308.

Romanes, G. J. (1978). *Cunningham's Manual of Practical Anatomy. Volume 3: Head and Neck and Brain.*
Oxford: University Press.

Rossi, M. (1971). *Le seuil de glissando ou seuil de perception des variations tonales pour les sons de la parole.*
Phonetica, vol. 23, pp. 1-33.

Rossi, M (1978). *La perception des glissandos descendants dans les contours prosodiques.*
Phonetica, vol. 35, pp. 11-40.

Samuel, A.G. (1977). *The Effect of Discrimination Training on Speech Perception: Noncategorical Perception.*
Perception and Psychophysics, vol. 22, pp. 321-320.

Samuel, A.G. (1982). *Phonetic Prototypes.*
Perception and Psychophysics, vol. 31, pp. 307-314.

Schiefer, L, Batliner, A. (1988). *Intonation, Ordnungseffekt und das Paradigma der kategorialen Wahrnehmung.*
IN Altmann (ed.), Intonationsforschungen. Tübingen: Max Niemeyer. pp. 273-291.

Seber, G. A. F. (1984). *Multivariate Observations.*
New York: John Wiley.

Shower, E.G., Biddulph, R. (1931). *Differential Pitch Sensitivity of the Ear.*
J. Acoust. Soc. Am., vol. 3. pp. 275-287.

Silverman, K. E. A., Pierrehumbert, J. B. (in press). *The Timing of Prenuclear High Accents in English.*
IN Kingston & Beckman (eds.), Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech. Cambridge: University Press.

Sonesson, B. (1974). *The Functional Anatomy of the Speech Organs.*
IN Malmberg (ed.), Manual of Phonetics. Amsterdam: North-Holland. pp. 45-75.

SPSS-X User's Guide. 3rd Edition. (1988).

Steele, S.A. (1986). *Interaction of Vowel F0 and Prosody.*
Phonetica, vol. 43, pp. 92-105.

Stevens, S.S., Volkmann, J., Newman, E.B. (1937). *A Scale for the Measurement of the Psychological Magnitude Pitch.*
J. Acoust. Soc. Am., vol. 8, pp. 185-190.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. & Cooper, F.S. (1970). *Motor Theory of Speech Perception: A Reply to Lane's Critical Review.*
Psychological Review, vol. 77, pp. 234-249.

Sundberg, J. (1979). *Maximum Speed of Pitch Changes in Singers and Untrained Subjects.*
J. of Phonetics, vol. 7, pp. 71-79.

Thorsen, N. (1988). *Standard Danish Intonation.*
ARIPUC, vol. 22, pp. 2-23.

Uselding, D.K. (1977). *A Temporal Order Effect in Voice Onset Time Discrimination.*
Language and Speech, vol. 20, pp. 366-377.

Verhoeven, J. (1986). *On the Relationship between Phonological Judgement and Perceived Prominence. A study with Reference to English.*
Univ. of Edinburgh, unpublished M. Sc. dissertation.

Verhoeven, J. (1987). *Categorical Perception and the Reality of Targets in Intonation. Two Pilot Experiments.*
Univ. of Edinburgh, Dep. of Linguistics, Work in Progress, vol. 20, pp. 74-89.

Verhoeven, J. (1988). *Order Effect in Discrimination Experiments on Dutch Intonation: towards Phonetic Prototypes.*
Paper presented at the Dep. of Linguistics Postgraduate Conference, University of Edinburgh, May 26th.

Wakita, H. (1980). *New Methods of Analysis in Speech Acoustics.*
Phonetica, vol. 37, pp. 87-108.

Wier, C.C., Yestead, W., Green, D.M. (1977). *Frequency Discrimination as a Function of Frequency and Sensation Level.*
J. Acoust. Soc. Am., vol. 61, pp. 178-184.

Willems, N.J. (1982). *English Intonation from a Dutch Point of View.*
Dordrecht: Foris.

Willems, N.J. (1983). *STEP: A model of Standard English Intonation.*
IPO Annual Progress Report, vol. 18, pp. 37-42.

Willems, N., Collier, R., 't Hart, J. (1988). *A Synthesis Scheme for British English Intonation.*
J.Acoust. Soc. Am., vol. 84, pp. 1250-1261.

Wood, C. C. (1976). *Discriminability, Response Bias and Phoneme Categories in Discrimination of VOT.*
J. Acoust. Soc. Am., vol. 59, pp. 811-812.

Woodrow, H. (1951). *Time Perception.*
IN Stevens (ed.), Handbook of Experimental Psychology. New York: Wiley & Sons, pp. 1224-1236.

Woodworth, H. S. & Schlosberg, R. (1954). *Experimental Psychology.*
No Publisher.

APPENDIX

Table A1 : Location of the rises in the different stimulus continua in terms of the ILS frame numbers in each stimulus file. Each frame is 2.5 msec. In order to obtain the correct stimulus number, the asterisk should be replaced by '1' for stimuli with rise duration 50 msec, '2' for rise duration 70 msec, '4' for rise duration 90 msec and '5' for stimuli with rise duration 110 msec.

	Rise duration			
Stimulus Number	50 msec	70 msec	90 msec	110 msec
*02	633-653	633-661	633-669	633-677
*03	637-657	637-665	637-673	637-681
*04	641-661	641-669	641-677	641-685
*05	645-665	645-673	645-681	645-689
*06	649-669	649-677	649-685	649-693
*07	653-673	653-681	653-689	653-697
*08	657-677	657-685	657-693	657-701
*09	661-681	661-689	661-697	661-705
*10	665-685	665-693	665-701	665-709
*11	669-689	669-697	669-705	669-713
*12	673-693	673-701	673-709	673-717
*13	677-697	677-705	677-713	677-721
*14	681-701	681-709	681-717	681-725
*15	685-705	685-713	685-721	685-729
*16	689-709	689-717	689-725	689-733
*17	693-713	693-721	693-729	693-737
*18	697-717	697-725	697-733	697-741
*19	701-721	701-729	701-737	701-745
*20	705-725	705-733	705-741	705-749
*21	709-729	709-737	709-745	709-753
*22	713-733	713-741	713-749	713-757
*23	717-737	717-745	717-753	717-761
*24	721-741	721-749	721-757	721-765
*25	725-745	725-753	725-761	725-769
*26	729-749	729-757	729-765	
*27	733-753	733-761	733-769	
*28	737-757	737-765		
*29	741-761	741-769		
*30	745-765			
*31	749-769			

Table A2 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 50 msec. Presentation order is AB. Total number of observations per item is 19.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
102x105	10	54	102x106	5	26	102x107	6	32	102x108	6	32	102x109	4	20
103x106	6	33	103x107	3	16	103x108	6	32	103x109	8	42	103x110	4	20
104x107	3	18	104x108	4	20	104x109	8	42	104x110	5	26	104x111	5	26
105x108	6	33	105x109	6	32	105x110	7	37	105x111	5	26	105x112	8	42
106x109	5	28	106x110	7	37	106x111	4	21	106x112	4	20	106x113	4	20
107x110	5	28	107x111	2	11	107x112	6	32	107x113	6	32	107x114	4	32
108x111	8	44	108x112	3	16	108x113	7	37	108x114	5	26	108x115	5	26
109x112	8	44	109x113	6	32	109x114	8	42	109x115	3	16	109x116	4	20
110x113	8	44	110x114	3	16	110x115	7	37	110x116	2	11	110x117	6	32
111x114	5	28	111x115	6	32	111X116	7	37	111X117	3	16	111X118	8	42
112x115	8	44	112x116	7	37	112x117	6	32	112x118	6	32	112x119	7	37
113x116	7	38	113x117	8	42	113x118	4	21	113x119	5	26	113x120	5	26
114x117	6	33	114x118	7	37	114x119	7	37	114x120	6	32	114x121	5	26
115x118	5	28	115x119	3	16	115x120	11	58	115x121	8	42	115x122	6	32
116x119	4	23	116x120	3	66	116x121	5	26	116x122	7	37	116x123	4	20
117x120	5	28	117x121	6	32	117x122	7	37	117x123	2	11	117x124	4	20
118x121	5	28	118x122	8	42	118x123	4	21	118x124	4	20	118x125	9	47
119x122	3	18	119x123	5	26	119x124	10	53	119x125	5	26	119x126	9	47
120x123	4	23	120x124	4	20	120x125	3	16	120x126	5	26	120x127	9	47
121x124	4	23	121x125	6	32	121x126	6	32	121x127	7	37	121x128	12	63
122x125	6	33	122x126	8	42	122x127	9	47	122x128	8	42	122x129	11	58
123x126	8	44	123x127	7	37	123x128	7	37	123x129	9	47	123x130	8	42
124x127	7	38	124x128	11	58	124x129	9	47	124x130	9	47	124x131	10	53
125x128	6	33	125x129	10	53	125x130	6	32	125x131	7	37			
126x129	6	33	126x130	11	58	126x131	11	58						
127x130	10	54	127x131	5	26									
128x131	9	46												

Table A3 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 70 msec. Presentation order is AB. Total number of observations per item is 19.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
202x205	6	32	202x206	9	47	202x207	4	21	202x208	5	26	202x209	6	32
203x206	12	63	203x207	6	32	203x208	7	37	203x209	5	26	203x210	7	37
204x207	6	32	204x208	7	37	204x209	7	37	204x210	5	26	204x211	7	37
205x208	7	37	205x209	5	26	205x210	2	11	205x211	6	32	205x212	3	16
206x209	8	42	206x210	4	21	206x211	4	21	206x212	9	47	206x213	3	16
207x210	5	26	207x211	3	16	207x212	6	32	207x213	5	26	207x214	5	26
208x211	3	16	208x212	5	26	208x213	3	16	208x214	2	11	208x215	7	37
209x212	2	11	209x213	5	26	209x214	7	37	209x215	5	26	209x216	4	21
210x213	5	26	210x214	6	32	210x215	5	26	210x216	8	42	210x217	8	42
211x214	4	20	211x215	8	42	211x216	6	32	211x217	4	21	211x218	3	16
212x215	5	26	212x216	6	32	212x217	5	26	212x218	4	21	212x219	4	21
213x216	4	20	213x217	7	37	213x218	3	16	213x219	8	42	213x220	2	11
214x217	5	26	214x218	9	47	214x219	10	53	214x220	2	11	214x221	3	16
215x218	4	20	215x219	7	37	215x220	5	26	215x221	4	21	215x222	3	16
216x219	3	16	216x220	7	37	216x221	5	26	216x222	6	32	216x223	6	32
217x220	4	20	217x221	5	26	217x222	9	47	217x223	6	32	217x224	4	21
218x221	7	37	218x222	5	26	218x223	4	21	218x224	4	21	218x225	8	42
219x222	6	32	219x223	3	16	219x224	7	37	219x225	6	32	219x226	9	47
220x223	7	37	220x224	5	26	220x225	1	11	220x226	7	37	220x227	10	53
221x224	8	42	221x225	6	32	221x226	8	42	221x227	8	42	221x228	9	47
222x225	3	16	222x226	8	42	222x227	7	37	222x228	9	47	222x229	10	53
223x226	9	47	223x227	8	42	223x228	7	37	223x229	9	47			
224x227	6	32	224x228	11	58	224x229	4	21						
225x228	7	37	225x229	10	53									
226x229	7	37												

Table A4 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 90 msec. Presentation order is AB. Total number of observations per itemis 25.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
402x405	8	32	402x406	7	28	402x407	4	16	402x408	3	12	402x409	3	12
403x406	4	16	403x407	4	16	403x408	3	12	403x409	5	20	403x410	4	16
404x407	8	32	404x408	7	28	404x409	5	20	404x410	7	28	404x411	6	24
405x408	7	28	405x409	6	24	405x410	4	16	405x411	4	16	405x412	4	16
406x409	9	36	406x410	7	28	406x411	3	12	406x412	5	20	406x413	5	20
407x410	3	12	407x411	6	24	407x412	1	4	407x413	5	20	407x414	9	36
408x411	6	24	408x412	8	32	408x413	5	20	408x414	5	20	408x415	5	20
409x412	7	28	409x413	6	24	409x414	8	32	409x415	3	12	409x416	3	12
410x413	8	32	410x414	6	24	410x415	8	32	410x416	6	24	410x417	9	36
411x414	7	28	411x415	5	20	411x416	9	36	411x417	3	12	411x418	5	20
412x415	5	20	412x416	8	32	412x417	7	28	412x418	5	20	412x419	7	28
413x416	8	32	413x417	5	20	413x418	3	12	413x419	5	20	413x420	8	32
414x417	8	32	414x418	7	28	414x419	7	28	414x420	5	20	414x421	10	40
415x418	7	28	415x419	11	44	415x420	9	36	415x421	5	20	415x422	8	32
416x419	7	28	416x420	6	24	416x421	8	32	416x422	6	24	416x423	6	24
417x420	7	28	417x421	9	36	417x422	7	28	417x423	7	28	417x424	11	44
418x421	10	40	418x422	7	28	418x423	8	32	418x424	10	40	418x425	7	28
419x422	12	48	419x423	8	32	419x424	4	16	419x425	9	24	419x426	8	32
420x423	9	36	420x424	8	32	420x425	11	44	420x426	9	24	420x427	6	24
421x424	8	32	421x425	11	44	421x426	12	48	421x427	7	28			
422x425	12	48	422x426	12	48	422x427	6	24						
423x426	9	36	423x427	7	28									
424x427	12	48												

Table A5: 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 110 msec. Presentation order is AB. Total number of observations per item is 25.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
502x505	7	28	502x506	12	48	MISSING	502x508	5	20	502x509	6	24
503x506	3	12	503x507	6	24		503x509	6	24	503x510	3	12
504x507	4	16	504x508	7	28		504x510	5	20	504x511	4	16
505x508	6	24	505x509	8	32		505x511	6	24	505x512	3	12
506x509	4	16	506x510	7	28		506x512	4	16	506x513	5	20
507x510	3	12	507x511	4	16		507x513	13	52	507x514	9	36
508x511	4	16	508x512	5	20		508x514	5	20	508x515	6	24
509x512	6	24	509x513	8	32		509x515	6	24	509x516	7	28
510x513	6	24	510x514	9	36		510x516	7	28	510x517	6	24
511x514	6	24	511x515	7	28		511x517	7	28	511x518	3	12
512x515	8	32	512x516	9	36		512x518	3	12	512x519	8	32
513x516	6	24	513x517	7	28		513x519	9	36	513x520	2	6
514x517	5	20	514x518	5	20		514x520	9	36	514x521	5	20
515x518	5	20	515x519	8	32		515x521	9	36	515x522	10	40
516x519	6	24	516x520	8	32		516x522	8	32	516x523	7	28
517x520	6	24	517x521	7	28		517x523	7	28	517x524	10	40
518x521	8	32	518x522	9	36		518x524	9	36	518x525	9	36
519x522	12	48	519x523	10	40		519x525	11	44			
520x523	8	32	520x524	6	24							
521x524	10	40	521x525	10	40							
522x525	11	44										

Table A6 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 50 msec. Presentation order is BA. Total number of observations per item is 22.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
105x102	12	55	106x102	10	45	107x102	8	36	108x102	10	45	109x102	11	50
106x103	10	45	107x103	5	23	108x103	10	45	109x103	7	32	110x103	12	55
107x104	9	41	108x104	12	55	109x104	10	45	110x104	13	59	111x104	13	59
108x105	10	45	109x105	13	59	110x105	11	50	111x105	11	50	112x105	14	64
109x106	15	68	110x106	11	50	111x106	10	45	112x106	13	59	113x106	12	55
110x107	14	64	111x107	11	50	112x107	8	36	113x107	8	36	114x107	14	64
111x108	7	33	112x108	7	32	113x108	10	45	114x108	12	55	115x108	10	45
112x109	9	41	113x109	14	64	114x109	11	50	115x109	10	45	116x109	14	64
113x110	13	59	114x110	9	41	115x110	11	50	116x110	10	45	117x110	10	45
114x111	10	45	115x111	11	50	116x111	10	45	117x111	9	41	118x111	12	55
115x112	13	59	116x112	10	45	117x112	9	41	118x112	10	45	119x112	7	32
116x113	10	45	117x113	4	18	118x113	6	27	119x113	7	32	120x113	9	41
117x114	3	14	118x114	7	32	119x114	8	36	120x114	6	27	121x114	9	41
118x115	13	59	119x115	4	18	120x115	9	41	121x115	9	41	122x115	9	41
119x116	6	27	120x116	5	23	121x116	4	18	122x116	6	27	123x116	12	55
120x117	4	36	121x117	9	41	122x117	7	32	123x117	6	27	124x117	8	36
121x118	6	27	122x118	6	27	123x118	6	27	124x118	6	27	125x118	5	28
122x119	7	32	123x119	10	45	124x119	9	41	125x119	10	45	126x119	12	55
123x120	9	41	124x120	5	23	125x120	6	27	126x120	5	23	127x120	13	59
124x121	6	27	125x121	4	18	126x121	5	28	127x121	9	41	128x121	9	41
125x122	3	14	126x122	5	23	127x122	5	28	128x122	6	27	129x122	6	27
126x123	6	27	127x123	4	18	128x123	7	32	129x123	7	32	130x123	7	32
127x124	3	14	128x124	6	27	129x124	5	28	130x124	9	41	131x124	7	32
128x125	7	32	129x125	3	14	130x125	5	28	131x125	6	27			
129x126	8	36	130x126	6	27	131x126	5	28						
130x127	7	32	131x127	5	28									
131x128	3	14												

Table A7 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 70 msec. Presentation order is BA. Total number of observations per item is 22.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
205x202	6	27	206x202	12	55	207x202	11	50	208x202	11	50	209x202	11	50
206x203	9	41	207x203	10	45	208x203	12	55	209x203	14	64	210x203	12	55
207x204	9	41	208x204	14	64	209x204	15	68	210x204	13	59	211x204	13	59
208x205	13	59	209x205	11	50	210x205	13	59	211x205	11	50	212x205	14	64
209x206	12	55	210x206	17	77	211x206	14	64	212x206	9	41	213x206	13	59
210x207	12	55	211x207	6	27	212x207	10	45	213x207	13	59	214x207	14	64
211x208	5	28	212x208	12	55	213x208	7	32	214x208	9	41	215x208	9	41
212x209	7	32	213x209	11	50	214x209	11	50	215x209	6	27	216x209	8	36
213x210	10	45	214x210	5	23	215x210	7	32	216x210	10	45	217x210	10	45
214x211	3	14	215x211	12	55	216x211	9	41	217x211	9	41	218x211	12	55
215x212	7	32	216x212	8	36	217x212	13	59	218x212	8	36	219x212	12	55
216x213	4	18	217x213	9	41	218x213	14	64	219x213	6	27	220x213	13	59
217x214	7	32	218x214	9	41	219x214	6	27	220x214	9	41	221x214	9	41
218x215	4	18	219x215	7	32	220x215	8	36	221x215	5	23	222x215	14	64
219x216	8	36	220x216	9	41	221x216	11	50	222x216	9	41	223x216	11	50
220x217	2	9	221x217	6	27	222x217	11	50	223x217	10	45	224x217	7	32
221x218	8	36	222x218	6	27	223x218	6	27	224x218	6	27	225x218	10	45
222x219	9	41	223x219	7	32	224x219	8	36	225x219	4	18	226x219	7	32
223x220	5	28	224x220	6	27	225x220	6	27	226x220	4	18	227x220	7	32
224x221	5	28	225x221	5	23	226x221	7	32	227x221	9	41	228x221	13	59
225x222	2	9	226x222	4	18	227x222	5	23	228x223	4	18	229x222	6	27
226x223	4	18	227x223	7	32	228x223	8	36	229x224	4	18			
227x224	2	9	228x224	5	23	229x224	5	23						
228x225	5	28	229x225	3	14									
229x226	5	28												

Table A8 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 90 msec. Presentation order is BA. Total number of observations per item is 30.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
405x402	14	47	406x402	17	57	407x402	10	33	408x402	10	33	409x402	12	40
406x403	14	47	407x403	14	47	408x403	14	47	409x403	15	50	410x403	15	50
407x405	19	63	408x404	15	50	409x404	14	47	410x404	11	37	411x404	15	50
408x406	14	47	409x405	9	30	410x405	11	37	411x405	14	47	412x405	14	47
409x407	18	60	410x406	6	20	411x406	16	53	412x406	13	43	413x406	8	27
410x408	5	17	411x407	10	33	412x407	11	37	413x407	11	37	414x407	17	57
411x409	11	37	412x408	11	37	413x408	12	40	414x408	17	57	415x408	3	10
412x410	11	37	413x409	11	37	414x409	11	37	415x409	14	47	416x409	11	37
413x411	8	27	414x410	12	40	415x410	14	47	416x410	12	40	417x410	10	33
414x412	9	30	415x411	7	23	416x411	12	40	417x411	13	43	418x411	9	30
415x413	5	17	416x412	7	23	417x412	8	27	418x412	12	40	419x412	9	30
416x414	11	37	417x413	11	37	418x413	7	23	419x413	7	23	420x413	12	40
417x415	4	13	418x414	4	13	419x414	11	37	420x414	8	27	421x414	10	33
418x416	4	13	419x415	6	20	420x415	7	23	421x415	8	27	422x415	10	33
419x417	5	17	420x416	10	33	421x416	6	20	422x416	12	40	423x416	15	50
420x418	8	27	421x417	6	20	422x417	9	30	423x417	10	23	424x417	9	30
421x419	6	20	422x418	7	23	423x418	7	23	424x418	14	47	425x418	6	20
422x420	8	27	423x419	3	10	424x419	7	23	425x419	9	30	426x419	7	23
423x421	5	17	424x420	8	27	425x420	5	17	426x420	9	30	427x420	11	37
424x422	6	20	425x421	5	17	426x421	7	23	427x421	5	17			
425x423	7	23	426x422	10	33	427x422	7	23						
426x424	13	43	427x423	8	27									
427x425	10	33												

Table A9 : 'Different' judgements for rise-alignment discrimination in the flat hat pattern for gap sizes 30 to 70 msec. Rise duration is 110 msec. Presentation order is BA. Total number of observations per item is 30.

ITEM NUMBER Gap 30 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 60 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 70 msec	TOT DIFF	% DIFF
505x502	10	33	506x502	12	40	507x502	12	40	508x502	13	43	509x502	17	57
506x503	10	33	507x503	17	57	508x503	12	40	509x503	10	33	510x503	9	30
507x504	16	53	508x504	12	40	509x504	13	43	510x504	14	47	511x504	12	40
508x505	8	27	509x505	6	20	510x505	15	50	511x505	13	43	512x505	15	50
509x506	4	13	510x506	13	43	511x506	10	33	512x506	8	27	513x506	18	60
510x507	11	37	511x507	17	57	512x507	10	33	513x507	6	20	514x507	11	37
511x508	4	13	512x508	14	47	513x508	14	47	514x508	11	37	515x508	15	50
512x509	10	33	513x509	9	30	514x509	7	23	515x509	8	27	516x509	15	50
513x510	9	30	514x510	9	30	515x510	12	40	516x510	9	30	517x510	10	33
514x511	3	10	515x511	7	23	516x511	6	20	517x511	8	27	518x511	11	37
515x512	4	13	516x512	10	33	517x512	8	27	518x512	11	37	519x512	12	40
516x513	4	13	517x513	10	33	518x513	8	27	519x513	8	27	520x513	9	30
517x514	3	10	518x514	7	23	519x514	7	23	520x514	15	50	521x514	9	30
518x515	5	17	519x515	13	43	520x515	7	23	521x515	6	20	522x515	8	27
519x516	4	33	520x516	10	33	521x516	9	30	522x516	10	33	523x516	11	37
520x517	8	27	521x517	10	33	522x517	8	27	523x517	7	23	524x517	13	43
521x518	13	43	522x518	6	20	523x518	9	30	524x518	10	33	525x518	9	30
522x519	6	20	523x519	11	37	524x519	-	-	525x519	8	27			
523x520	9	30	524x520	5	17	525x520	6	20						
524x521	5	17	525x521	6	20									
525x522	10	33												

Table A10: Number and percentage of 'high' judgements for rise-alignment in the labelling task of chapter 4. The rise-endpoint is given in ILS frame numbers and the total number of observations is 11,076.

RISE ENDPOINT	NUMBER LOW	NUMBER HIGH	% HIGH
653	8	70	90
657	15	63	81
661	24	132	85
665	24	132	85
669	42	191	82
673	36	199	85
677	69	243	78
681	44	268	86
685	67	243	78
689	56	256	82
693	72	240	77
697	78	234	75
701	85	227	73
705	83	229	73
709	105	206	66
713	113	199	64
717	129	182	58
721	133	179	57
725	145	166	53
729	149	163	52
733	161	151	48
737	210	102	33
741	210	102	33
745	215	97	31
749	239	71	23
753	242	70	22
757	239	73	23
761	256	55	18
765	262	50	16
769	244	67	22

Table A11 : Subject response coincidence matrix for rise-alignment discrimination in the flat hat pattern. Rise duration is 50 and 70 msec. Presentation order is AB. Number of subjects is 19. The dissimilarity metric ranges from zero for perfect correspondence to one for completely different pairings.

	SUB 1	SUB 2	SUB 3	SUB 4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11	SUB 12	SUB 13	SUB 14	SUB 15	SUB 16	SUB 17	SUB 18	SUB 19
SUB 1	0.0000																		
SUB 2	0.6209	0.0000																	
SUB 3	0.6102	0.5919	0.0000																
SUB 4	0.6236	0.6414	0.6261	0.0000															
SUB 5	0.7024	0.6122	0.6010	0.6920	0.0000														
SUB 6	0.6117	0.6402	0.6248	0.5771	0.6548	0.0000													
SUB 7	0.6664	0.5724	0.6280	0.6443	0.4997	0.6514	0.0000												
SUB 8	0.6337	0.6168	0.6124	0.6480	0.5510	0.6579	0.5386	0.0000											
SUB 9	0.6784	0.5750	0.6181	0.6755	0.5296	0.6372	0.5329	0.5821	0.0000										
SUB 10	0.7167	0.5775	0.5705	0.7258	0.4931	0.6524	0.5507	0.5724	0.4649	0.0000									
SUB 11	0.7306	0.5463	0.6021	0.7614	0.4017	0.6705	0.4911	0.5873	0.4557	0.3633	0.0000								
SUB 12	0.6620	0.6069	0.5824	0.6832	0.5211	0.6474	0.5341	0.5517	0.5173	0.4630	0.4209	0.0000							
SUB 13	0.6015	0.5931	0.6258	0.6356	0.6114	0.5873	0.6171	0.6085	0.6005	0.6090	0.6367	0.6046	0.0000						
SUB 14	0.6592	0.5944	0.6146	0.6352	0.5899	0.6237	0.5948	0.5996	0.5865	0.6036	0.6016	0.6222	0.6428	0.0000					
SUB 15	0.6622	0.5504	0.5607	0.6658	0.5415	0.5998	0.5642	0.5899	0.5362	0.4973	0.4432	0.5216	0.5836	0.5628	0.0000				
SUB 16	0.6305	0.5917	0.6201	0.6684	0.6342	0.6459	0.6356	0.6122	0.6010	0.6231	0.6531	0.6222	0.6099	0.6039	0.6367	0.0000			
SUB 17	0.6919	0.5766	0.6091	0.7294	0.4626	0.6681	0.5012	0.5418	0.5139	0.4324	0.3564	0.4592	0.6029	0.5842	0.4674	0.6501	0.0000		
SUB 18	0.5996	0.5210	0.5211	0.6274	0.5692	0.5940	0.6101	0.6097	0.5915	0.5832	0.5621	0.5337	0.5838	0.5857	0.5255	0.6190	0.5825	0.0000	
SUB 19	0.6517	0.5873	0.5956	0.6953	0.5358	0.6322	0.5885	0.6095	0.5647	0.5248	0.4966	0.6015	0.6090	0.5798	0.5401	0.5862	0.5143	0.5910	0.0000

Table A12 : Subject response coincidence matrix for rise-alignment discimination in the flat hat pattern. Rise duration is 50 and 70 msec. Presentation order is BA. Number of subjects is 22. The dissimilarity metric ranges from zero for perfect correspondence to one for completely different pairings.

	SUB 1	SUB 2	SUB 3	SUB 4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11	SUB 12	SUB 13	SUB 14	SUB 15	SUB 16	SUB 17	SUB 18	SUB 19	SUB 20	SUB 21	SUB 22
SUB 1	0.0000																					
SUB 2	0.5132	0.0000																				
SUB 3	0.5324	0.4425	0.0000																			
SUB 4	0.5881	0.6047	0.6041	0.0000																		
SUB 5	0.6152	0.6100	0.6143	0.5856	0.0000																	
SUB 6	0.7961	0.8150	0.7987	0.7277	0.7427	0.0000																
SUB 7	0.6266	0.6401	0.6245	0.6281	0.6409	0.5919	0.0000															
SUB 8	0.5253	0.4793	0.5478	0.5922	0.6142	0.8183	0.6182	0.0000														
SUB 9	0.6168	0.5942	0.6539	0.6111	0.6257	0.6283	0.5798	0.6253	0.0000													
SUB 10	0.8367	0.8499	0.8311	0.7574	0.7605	0.4031	0.6641	0.8575	0.6542	0.0000												
SUB 11	0.7983	0.8061	0.7940	0.7207	0.7275	0.4518	0.6113	0.8137	0.6387	0.3544	0.0000											
SUB 12	0.7571	0.7599	0.7480	0.6962	0.6962	0.4606	0.6503	0.7774	0.6369	0.4718	0.4911	0.0000										
SUB 13	0.6820	0.6828	0.6707	0.6588	0.6509	0.5370	0.5352	0.6707	0.5778	0.5829	0.6003	0.5795	0.0000									
SUB 14	0.7388	0.7514	0.7336	0.6860	0.6827	0.5132	0.5934	0.7545	0.6323	0.5091	0.5538	0.5368	0.5530	0.0000								
SUB 15	0.8091	0.8167	0.8015	0.7406	0.7377	0.4421	0.6488	0.8264	0.6584	0.3648	0.4275	0.5094	0.5849	0.5179	0.0000							
SUB 16	0.7457	0.7469	0.7357	0.6911	0.6998	0.5112	0.6367	0.7635	0.6306	0.4887	0.5139	0.5507	0.5865	0.5674	0.5238	0.0000						
SUB 17	0.7335	0.7347	0.7135	0.6848	0.6998	0.5158	0.5665	0.7450	0.6175	0.4849	0.4757	0.5401	0.4957	0.5362	0.5203	0.5434	0.0000					
SUB 18	0.7602	0.7724	0.7416	0.7133	0.7052	0.4548	0.5959	0.7775	0.6222	0.4746	0.5270	0.5130	0.5359	0.4935	0.4957	0.5329	0.5137	0.0000				
SUB 19	0.7203	0.6876	0.7006	0.6592	0.6592	0.5652	0.6047	0.7362	0.5957	0.4912	0.4516	0.5265	0.5869	0.6193	0.5249	0.5647	0.5545	0.6073	0.0000			
SUB 20	0.6896	0.6736	0.6707	0.6691	0.6798	0.5123	0.5567	0.6805	0.5899	0.6029	0.5881	0.5795	0.5089	0.5386	0.5914	0.5865	0.5124	0.5282	0.6181	0.0000		
SUB 21	0.5951	0.5959	0.5746	0.6259	0.6122	0.6820	0.5317	0.5954	0.5659	0.7171	0.6707	0.6622	0.5759	0.6588	0.7081	0.6315	0.6331	0.6177	0.6094	0.5824	0.0000	
SUB 22	0.7370	0.7565	0.7409	0.6625	0.6716	0.5066	0.5937	0.7660	0.6241	0.5098	0.5010	0.5471	0.5649	0.5829	0.5193	0.5573	0.5607	0.5644	0.5098	0.6011	0.6448	0.0000

	SUB 1	SUB 2	SUB 3	SUB4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11	SUB 12	SUB 13	SUB 14	SUB 15	SUB 16	SUB 17	SUB 18	SUB 19	SUB 20	SUB 21	SUB 22	SUB 23	SUB 24	SUB 25
SUB 1	0.0000																								
SUB 2	0.1875	0.0000																							
SUB 3	0.5030	0.4979	0.0000																						
SUB 4	0.8295	0.8804	0.7184	0.0000																					
SUB 5	0.6919	0.6983	0.6306	0.6691	0.0000																				
SUB 6	0.5234	0.4948	0.5592	0.7373	0.5980	0.0000																			
SUB 7	0.7096	0.7264	0.6475	0.6382	0.5996	0.6189	0.0000																		
SUB 8	0.3939	0.2938	0.5294	0.7799	0.6333	0.528	0.6575	0.0000																	
SUB 9	0.4135	0.3831	0.5436	0.7649	0.6261	0.5522	0.6791	0.4656	0.0000																
SUB 10	0.4729	0.3917	0.5746	0.7600	0.6473	0.5276	0.6554	0.4839	0.5454	0.0000															
SUB 11	0.7317	0.7374	0.6435	0.5980	0.6335	0.6325	0.6026	0.6842	0.6722	0.6225	0.0000														
SUB 12	0.3705	0.2644	0.5764	0.7986	0.6453	0.4436	0.6411	0.3773	0.4488	0.4917	0.6659	0.0000													
SUB 13	0.4708	0.4475	0.5928	0.7320	0.6384	0.5313	0.6261	0.4428	0.5411	0.5550	0.6466	0.4258	0.0000												
SUB 14	0.4346	0.4063	0.5572	0.7617	0.6283	0.5172	0.6610	0.4839	0.5454	0.4991	0.6860	0.4800	0.5550	0.0000											
SUB 15	0.6596	0.6584	0.6331	0.6439	0.5993	0.5945	0.5694	0.5759	0.6346	0.6365	0.5828	0.5584	0.5835	0.6165	0.0000										
SUB 16	0.6301	0.6409	0.5653	0.6610	0.6195	0.5992	0.6467	0.5975	0.5769	0.6122	0.6315	0.6323	0.6038	0.6192	0.6188	0.0000									
SUB 17	0.6988	0.7201	0.6363	0.6538	0.6482	0.6496	0.6185	0.6864	0.6195	0.6494	0.6324	0.6701	0.6395	0.6796	0.6074	0.6089	0.0000								
SUB 18	0.5344	0.5796	0.6233	0.6993	0.6367	0.6051	0.6236	0.5909	0.5898	0.5850	0.6498	0.5924	0.5602	0.5589	0.5980	0.5967	0.6296	0.0000							
SUB 19	0.3467	0.3894	0.5670	0.7544	0.6607	0.5746	0.6826	0.4336	0.4820	0.4762	0.6762	0.5021	0.5265	0.4991	0.6389	0.6343	0.6639	0.5204	0.0000						
SUB 20	0.2826	0.1584	0.5322	0.8300	0.6727	0.5405	0.6855	0.3576	0.4072	0.4675	0.7186	0.3327	0.4527	0.4675	0.6508	0.6434	0.6786	0.5626	0.3850	0.0000					
SUB 21	0.5053	0.5010	0.5762	0.7106	0.6498	0.5844	0.6505	0.5204	0.5630	0.5080	0.6213	0.5694	0.5777	0.5989	0.6198	0.6040	0.6594	0.5893	0.5401	0.5451	0.0000				
SUB 22	0.4789	0.4979	0.5206	0.7012	0.6363	0.5842	0.6475	0.5188	0.5621	0.5661	0.6365	0.5386	0.5510	0.5661	0.6079	0.5903	0.6614	0.5472	0.5581	0.5212	0.5762	0.0000			
SUB 23	0.3711	0.2830	0.5265	0.8004	0.6419	0.5142	0.6663	0.4325	0.4605	0.4671	0.6753	0.3546	0.4635	0.4671	0.6279	0.6141	0.6726	0.5722	0.4539	0.3333	0.5486	0.5469	0.0000		
SUB 24	0.2828	0.2168	0.5055	0.8192	0.6580	0.4798	0.6895	0.3561	0.4184	0.4888	0.6866	0.3318	0.3799	0.4521	0.6017	0.6209	0.6834	0.5464	0.3969	0.2912	0.5190	0.4939	0.3477	0.0000	
SUB 25	0.2230	0.0859	0.5159	0.8653	0.6888	0.4890	0.7099	0.3074	0.4080	0.3873	0.7207	0.2791	0.4287	0.4161	0.6576	0.6409	0.7162	0.5824	0.3703	0.2136	0.5303	0.5041	0.2971	0.2150	0.0000

Table A14 : Subject coincidence response matrix for rise-alignment discrimination in the flat hat pattern. Rise duration is 90 and 110 msec. Presentation order is BA. Number of subjects is 30. The dissimilarity metric ranges from zero for perfect correspondence to one for completely different pairings.

	SUB 1	SUB 2	SUB 3	SUB 4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11	SUB 12	SUB 13	SUB 14	SUB 15	SUB 16	SUB 17	SUB 18	SUB 19	SUB 20	SUB 21	SUB 22	SUB 23	SUB 24	SUB 25	SUB 26	SUB 27	SUB 28	SUB 29	SUB 30
SUB 1	0.0000																													
SUB 2	0.4751	0.0000																												
SUB 3	0.5979	0.6141	0.0000																											
SUB 4	0.7256	0.7796	0.6669	0.0000																										
SUB 5	0.6073	0.6400	0.6060	0.6349	0.0000																									
SUB 6	0.5881	0.5942	0.6202	0.6818	0.5998	0.0000																								
SUB 7	0.5854	0.4843	0.6347	0.7671	0.6609	0.6243	0.0000																							
SUB 8	0.5707	0.5930	0.6354	0.6944	0.6433	0.5621	0.6335	0.0000																						
SUB 9	0.5348	0.5459	0.6393	0.7778	0.6766	0.5867	0.5670	0.5461	0.0000																					
SUB 10	0.6003	0.6323	0.6215	0.6933	0.6556	0.6151	0.6358	0.5782	0.5820	0.0000																				
SUB 11	0.6644	0.7076	0.6419	0.5924	0.6198	0.6354	0.6806	0.6186	0.6595	0.6417	0.0000																			
SUB 12	0.6419	0.5964	0.6368	0.7193	0.6215	0.5989	0.6019	0.6207	0.6162	0.6139	0.6348	0.0000																		
SUB 13	0.6142	0.6489	0.6080	0.6233	0.6112	0.6323	0.6731	0.6442	0.6811	0.6526	0.6293	0.6366	0.0000																	
SUB14	0.5113	0.4130	0.6679	0.8633	0.7256	0.6420	0.5054	0.5995	0.4726	0.5913	0.7575	0.6458	0.7567	0.0000																
SUB 15	0.5664	0.5874	0.6200	0.7330	0.6552	0.6378	0.5958	0.5817	0.5727	0.6354	0.6400	0.6197	0.6358	0.5561	0.0000															
SUB 16	0.5330	0.4339	0.6058	0.7521	0.6416	0.5877	0.5415	0.6093	0.6056	0.5989	0.6979	0.6046	0.6308	0.5010	0.5854	0.0000														
SUB 17	0.4758	0.3747	0.6488	0.8367	0.6958	0.5679	0.4822	0.5777	0.4940	0.6335	0.7363	0.6302	0.6969	0.3501	0.4822	0.4430	0.0000													
SUB 18	0.5382	0.4355	0.7129	0.9132	0.7587	0.6631	0.5086	0.5676	0.4599	0.6141	0.8019	0.6357	0.7912	0.2480	0.5548	0.5045	0.3071	0.0000												
SUB 19	0.5561	0.4620	0.6293	0.7722	0.6219	0.6172	0.4996	0.6310	0.5814	0.6291	0.7025	0.6184	0.6773	0.4677	0.5605	0.5048	0.4446	0.4670	0.0000											
SUB 20	0.5848	0.5047	0.6964	0.8633	0.7298	0.6350	0.5256	0.5916	0.4726	0.6219	0.7433	0.6519	0.7679	0.3279	0.5650	0.5408	0.3905	0.2638	0.4789	0.0000										
SUB 21	0.5554	0.4834	0.6829	0.8334	0.7007	0.6247	0.5327	0.5694	0.4940	0.6273	0.7017	0.6095	0.7326	0.4033	0.5238	0.5191	0.3696	0.3220	0.4893	0.3773	0.0000									
SUB 22	0.5491	0.4838	0.6561	0.7890	0.6411	0.6271	0.5384	0.5685	0.5585	0.6416	0.7113	0.5952	0.7029	0.4975	0.5621	0.5431	0.4656	0.4505	0.5190	0.4761	0.4766	0.0000								
SUB 23	0.5778	0.6165	0.6222	0.7385	0.6564	0.6192	0.5551	0.5190	0.5530	0.6060	0.6165	0.6106	0.6537	0.5870	0.5415	0.5958	0.5571	0.5612	0.5943	0.5789	0.5295	0.5891	0.0000							
SUB 24	0.5594	0.5722	0.5796	0.6963	0.6423	0.6079	0.5935	0.5943	0.6155	0.6115	0.6059	0.6017	0.6416	0.6265	0.6237	0.6157	0.6308	0.6676	0.6177	0.6265	0.6020	0.6014	0.6026	0.0000						
SUB 25	0.5930	0.5907	0.6099	0.7059	0.6083	0.6220	0.6096	0.6099	0.6237	0.5820	0.6383	0.6048	0.6400	0.6136	0.5960	0.5992	0.6193	0.6621	0.6006	0.6608	0.6121	0.6040	0.6114	0.5884	0.0000					
SUB 26	0.5431	0.5354	0.6134	0.6876	0.6366	0.6182	0.5850	0.6296	0.5992	0.6244	0.6722	0.6112	0.6126	0.5940	0.5790	0.5220	0.5704	0.6080	0.5207	0.6308	0.5948	0.5594	0.5962	0.5895	0.5785	0.0000				
SUB 27	0.5612	0.4741	0.7289	0.9179	0.7806	0.6745	0.5113	0.5896	0.4623	0.6177	0.8012	0.6396	0.8085	0.2489	0.5377	0.5287	0.3233	0.0970	0.4811	0.2327	0.3233	0.4280	0.5643	0.6646	0.6662	0.6282	0.0000			
SUB 28	0.5354	0.5250	0.5605	0.8119	0.6914	0.6573	0.5033	0.6278	0.5599	0.5789	0.7322	0.6291	0.7019	0.4372	0.5580	0.4665	0.4400	0.3919	0.5010	0.4604	0.4740	0.5196	0.5863	0.6230	0.6312	0.5896	0.3937	0.0000		
SUB 29	0.5940	0.5996	0.5679	0.6757	0.5796	0.6108	0.5833	0.6272	0.6131	0.6303	0.6576	0.6166	0.6229	0.6581	0.6098	0.5710	0.6147	0.6590	0.5802	0.6317	0.6417	0.5675	0.6182	0.6136	0.6477	0.5931	0.6772	0.5672	0.0000	
SUB 30	0.5342	0.4847	0.6534	0.8045	0.6881	0.6335	0.5404	0.5724	0.5426	0.6311	0.7099	0.5777	0.6983	0.4726	0.5651	0.5271	0.4164	0.3802	0.5384	0.4499	0.4522	0.5096	0.5354	0.6208	0.6405	0.5722	0.4088	0.5092	0.5650	0.0000

Table A15 : Fall location in the stimuli for the early, ambiguous and late rise conditions in terms of the ILS frame numbers. Each frame is 10 msec.

Stimulus Number	Fall Location
*00	168-176
*01	169-177
*02	170-178
*03	171-179
*04	172-180
*05	173-181
*06	174-182
*07	175-183
*08	176-184
*09	177-185
*10	178-186
*11	179-187
*12	180-188
*13	181-189
*14	182-190
*15	183-191
*16	184-192
*17	185-193
*18	186-194
*19	187-195
*20	188-196
*21	189-197
*22	190-198

Table A16: 'Different' judgements for fall-alignment discrimination in the flat hat pattern for gap sizes 40 and 50 msec. The rise location is ambiguous and the stimuli are presented in AB-presentation order. Total number of observations for each item is 55, except for the last four items in gap size 40 where the number of observation is 110.

ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF
400x404	20	36	400x405	20	36
401x405	16	29	401x406	17	31
402x406	22	40	402x407	20	36
403x407	22	40	403x408	26	47
404x408	25	45	404x409	20	36
405x409	30	55	405x410	33	60
406x410	15	27	406x411	26	47
407x411	27	49	407x412	33	60
408x412	19	35	408x413	28	51
409x413	30	55	409x414	37	67
410x414	33	60	410x415	39	71
411x415	28	51	411x416	39	71
412x416	28	51	412x417	28	51
413x417	28	51	413x418	28	51
414x418	29	53			
415x419	51	46			
416x420	40	36			
417x421	35	32			
418x422	37	34			

Table A17: 'Different' judgements for fall-alignment discrimination in the flat hat pattern for gap sizes 40 and 50 msec. The rise location is early and the stimuli are presented in AB-presentation order. Total number of observations for each item is 100, except for the last four items in gap size 40 where the number of observation is 200.

ITEM NUMBER Gap 40 msec	TOT DIFF	% DIFF	ITEM NUMBER Gap 50 msec	TOT DIFF	% DIFF
500x504	41	41	500x505	31	31
501x505	37	37	501x506	42	42
502x506	33	33	502x507	35	35
503x507	43	43	503x508	47	47
504x508	46	46	504x509	36	36
505x509	38	38	505x510	52	52
506x510	39	39	506x511	41	41
507x511	39	39	507x512	44	44
508x512	37	37	508x513	62	62
509x513	42	42	509x514	53	53
510x514	51	51	510x515	48	48
511x515	51	51	511x516	51	51
512x516	52	52	512x517	60	60
513x517	48	48	513x518	58	58
514x518	45	45			
515x519	96	48			
516x520	79	39			
517x521	70	35			
518x522	66	33			

Table A18 : Subject response coincidence matrix for fall-alignment discrimination in the hat pattern. Rise location is ambiguous. Presentation order is AB. Number of subjects is 11. The dissimilarity metric ranges from zero for perfect correspondence to one for completely different pairings.

	SUB 1	SUB 2	SUB 3	SUB 4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11
SUB 1	0.0000										
SUB 2	0.6216	0.0000									
SUB 3	0.5901	0.6471	0.0000								
SUB 4	0.5886	0.6456	0.6231	0.0000							
SUB 5	0.5901	0.5901	0.5886	0.5901	0.0000						
SUB 6	0.5456	0.5465	0.4970	0.5886	0.4369	0.0000					
SUB 7	0.6096	0.6577	0.5781	0.5721	0.5781	0.3318	0.0000				
SUB 8	0.6622	0.6952	0.7177	0.7192	0.7177	0.6622	0.6141	0.0000			
SUB 9	0.6847	0.6847	0.6517	0.6366	0.5676	0.5526	0.5090	0.5315	0.0000		
SUB 10	0.6832	0.6592	0.7177	0.6832	0.7207	0.7042	0.6246	0.5015	0.6351	0.0000	
SUB 11	0.6216	0.6216	0.6231	0.6456	0.6231	0.4955	0.5721	0.6952	0.6847	0.6982	0.0000

Table A19 : Subject response coincidence matrix for fall-alignment discimination in the hat pattern. Rise is located early. Presentation order is AB. Number of subjects is 20. The dissimilarity metric ranges from zero for perfect correspondence to one for completely different pairings.

	SUB 1	SUB 2	SUB 3	SUB 4	SUB 5	SUB 6	SUB 7	SUB 8	SUB 9	SUB 10	SUB 11	SUB 12	SUB 13	SUB 14	SUB 15	SUB 16	SUB 17	SUB 18	SUB 19	SUB 20
SUB 1	0.0000																			
SUB 2	0.6686	0.0000																		
SUB 3	0.6080	0.6269	0.0000																	
SUB 4	0.6837	0.6553	0.6686	0.0000																
SUB 5	0.6932	0.6648	0.6761	0.2254	0.0000															
SUB 6	0.6496	0.6932	0.6364	0.5909	0.4905	0.0000														
SUB 7	0.5417	0.6799	0.6364	0.6648	0.6288	0.5360	0.0000													
SUB 8	0.7273	0.5852	0.7045	0.4905	0.4811	0.6307	0.6515	0.0000												
SUB 9	0.8163	0.6136	0.7727	0.6515	0.5814	0.7841	0.7405	0.4905	0.0000											
SUB 10	0.6610	0.6894	0.4735	0.6610	0.6780	0.6042	0.6042	0.6742	0.7727	0.0000										
SUB 11	0.6610	0.6799	0.5341	0.6610	0.6780	0.6042	0.6572	0.6742	0.7708	0.5341	0.0000									
SUB 12	0.6136	0.8201	0.6174	0.7140	0.7424	0.5814	0.6515	0.8485	0.9602	0.6610	0.6610	0.0000								
SUB 13	0.6402	0.6894	0.6477	0.6610	0.6686	0.6364	0.6364	0.6742	0.7102	0.6212	0.5833	0.6610	0.0000							
SUB 14	0.6705	0.6269	0.6477	0.5644	0.5720	0.5606	0.6364	0.5436	0.7424	0.6212	0.6629	0.6610	0.6212	0.0000						
SUB 15	0.8371	0.6420	0.7689	0.7367	0.6932	0.8011	0.7424	0.5284	0.2727	0.7424	0.7689	0.9356	0.7424	0.7689	0.0000					
SUB 16	0.6951	0.6477	0.6894	0.6288	0.5852	0.6515	0.6004	0.5360	0.5038	0.6269	0.6799	0.8163	0.6269	0.6591	0.5436	0.0000				
SUB 17	0.8561	0.5814	0.7727	0.7140	0.6515	0.8011	0.7689	0.5038	0.3201	0.7727	0.7727	0.9811	0.7936	0.7936	0.2727	0.5814	0.0000			
SUB 18	0.7633	0.5720	0.7159	0.6951	0.6591	0.7367	0.7140	0.6080	0.5095	0.7311	0.7311	0.8769	0.7159	0.6894	0.4091	0.6477	0.4508	0.0000		
SUB 19	0.6496	0.6894	0.6364	0.5909	0.4905	0.3182	0.4754	0.5814	0.7841	0.6042	0.6042	0.5814	0.6042	0.5606	0.8030	0.6136	0.8011	0.7367	0.0000	
SUB 20	0.8011	0.5284	0.7557	0.6174	0.5436	0.7102	0.7159	0.2841	0.2727	0.7405	0.7405	0.9470	0.7140	0.6761	0.4034	0.5284	0.2756	0.5360	0.7100	0.0000

Table A20 : Pitch movement location in the stimuli for the JND of alignment in terms of ILS frame numbers. Each frame is 2.5 msec.

Stimulus Number	Rise Location	Stimulus Number	Fall Location
101	330-370	201	708-748
102	334-374	202	712-752
103	338-378	203	716-756
104	342-382	204	720-760
105	346-386	205	724-764
106	350-390	206	728-768
107	354-394	207	732-772
108	358-398	208	736-776
109	362-402	209	740-780
110	366-406	210	744-784
111	370-410	211	748-788
112	374-414	212	752-792
113	378-418	213	756-796
114	382-422	214	760-800
115	386-426	215	764-804
116	390-430	216	768-808
117	394-434	217	772-812
118	398-438	218	776-816
119	402-442	219	780-820

Table A21: 'Different' judgements for pitch movement alignment discrimination in the hat pattern. Presentation order is AB and the total number of observations for each item is 70.

ITEM NUMBER **Rise**	TOT DIFF	% DIFF	ITEM NUMBER **Fall**	TOT DIFF	% DIFF
101 x 102	13	19	201 x 202	1	1
101 x 103	13	19	201 x 203	9	13
101 x 104	9	13	201 x 204	13	19
101 x 105	15	21	201 x 205	14	20
101 x 106	15	21	201 x 206	27	39
101 x 107	20	29	201 x 207	27	39
101 x 108	18	26	201 x 208	39	56
101 x 109	23	33	201 x 209	41	59
101 x 110	29	41	201 x 210	46	66
101 x 111	29	41	201 x 211	51	73
101 x 112	38	54	201 x 212	65	93
101 x 113	44	63	201 x 213	60	86
101 x 114	48	69	201 x 214	62	89
101 x 115	44	63	201 x 215	67	96
101 x 116	52	74	201 x 216	66	94
101 x 117	58	83	201 x 217	68	97
101 x 118	66	94	201 x 218	70	100
101 x 119	51	73	201 x 219	68	97

Table A22 : Pitch movement location in the stimuli for the JND of alignment in terms of ILS frame numbers. Each frame number is 2.5 msec.

Stimulus Number	**Rise Location**	**Stimulus Number**	**Fall Location**
101	132 x 172	201	441x481
102	136 x 176	202	445x485
103	140 x 180	203	449x489
104	144 x 184	204	453x493
105	148 x 188	205	457x597
106	152 x 192	206	461x501
107	156 x 196	207	465x505
108	160 x 200	208	469x509
109	164 x 204	209	573x513
110	168 x 208	210	477x517
111	172 x 212	211	481x421
112	176 x 216	212	485x525
113	180 x 220	213	489x529
114	184 x 224	214	493x533
115	188 x 228	215	497x537
116	192 x 232	216	501x541
117	196 x 236	217	505x545
118	200 x 240	218	509x549
119	204 x 244	219	513x553
120	208 x 248	220	517x557
121	212 x 252	221	521x561

Table A23: 'Different' judgements for rise alignment discrimination in the hat pattern with an early rise in the standard stimulus. Total number of observations for each item is 55.

Early standard AB			**Early standard BA**		
ITEM NUMBER	TOT DIFF	% DIFF	ITEM NUMBER	TOT DIFF	% DIFF
101 x 102	6	11	102 x 101	3	6
101 x 103	6	11	103 x 101	7	14
101 x 104	8	15	104 x 101	4	8
101 x 105	11	20	105 x 101	11	22
101 x 106	8	15	106 x 101	6	12
101 x 107	11	20	107 x 101	10	20
101 x 108	7	13	108 x 101	8	16
101 x 109	10	18	109 x 101	13	26
101 x 110	16	29	110 x 101	11	22
101 x 111	17	31	111 x 101	17	34
101 x 112	21	38	112 x 101	22	44
101 x 113	30	55	113 x 101	17	34
101 x 114	40	73	114 x 101	23	46
101 x 115	41	75	115 x 101	24	48
101 x 116	44	80	116 x 101	25	50

Table A24: 'Different' judgements for rise alignment discrimination in the hat pattern with late rise in the standard stimulus. Total number of observations for each item is 55.

Late standard AB			**Late standard BA**		
ITEM NUMBER	TOT DIFF	% DIFF	ITEM NUMBER	TOT DIFF	% DIFF
106 x 107	3	5	107 x 106	5	10
106 x 108	4	7	108 x 106	4	8
106 x 109	6	11	109 x 106	5	10
106 x 110	7	13	110 x 106	7	14
106 x 111	6	11	111 x 106	6	12
106 x 112	5	9	112 x 106	8	16
106 x 113	8	15	113 x 106	10	20
106 x 114	14	25	114 x 106	10	20
106 x 115	25	45	115 x 106	17	34
106 x 116	24	44	116 x 106	15	30
106 x 117	30	55	117 x 106	16	32
106 x 118	38	69	118 x 106	23	46
106 x 119	37	67	119 x 106	20	40
106 x 120	39	71	120 x 106	29	58
106 x 121	43	78	121 x 106	25	50

Table A25: 'Different' judgements for fall alignment discrimination in the hat pattern with early fall in the standard stimulus. Total number of observations for each item is 55.

Late standard AB			**Late standard BA**		
ITEM NUMBER	TOT DIFF	% DIFF	ITEM NUMBER	TOT DIFF	% DIFF
201 x 202	13	26	202 x 201	15	27
201 x 203	7	14	203 x 201	21	38
201 x 204	10	20	204 x 201	10	18
201 x 205	6	12	205 x 201	11	20
201 x 206	8	16	206 x 201	16	29
201 x 207	11	22	207 x 201	17	31
201 x 208	24	48	208 x 201	30	55
201 x 209	23	46	209 x 201	29	53
201 x 210	27	54	210 x 201	31	56
201 x 211	36	72	211 x 201	28	51
201 x 212	39	78	212 x 201	36	65
201 x 213	39	78	213 x 201	33	60
201 x 214	43	86	214 x 201	37	67
201 x 215	47	94	215 x 201	36	65
201 x 216	45	90	216 x 201	48	87

Table A26: 'Different' judgements for fall alignment discrimination in the hat pattern with late fall in the standard stimulus. Total number of observations for each item is 55.

Late standard AB			**Late standard BA**		
ITEM NUMBER	TOT DIFF	% DIFF	ITEM NUMBER	TOT DIFF	% DIFF
206 x 207	11	22	207 x 206	10	20
206 x 208	8	16	208 x 206	7	14
206 x 209	13	26	209 x 206	18	36
206 x 210	17	34	210 x 206	17	34
206 x 211	21	42	211 x 206	23	46
206 x 212	31	62	212 x 206	22	44
206 x 213	33	66	213 x 206	28	56
206 x 214	32	64	214 x 206	33	66
206 x 215	45	90	215 x 206	34	68
206 x 216	37	74	216 x 206	37	74
206 x 217	43	86	217 x 206	38	76
206 x 218	43	86	218 x 206	39	78
206 x 219	46	92	219 x 206	34	68
206 x 220	45	90	220 x 206	34	68
206 x 221	46	92	221 x 206	37	74

Table A27: 'Different' judgements for rise alignment discrimination in the hat pattern with gap size 110 msec. Total number of observations for each item is 50.

AB Order			**BA Order**		
ITEM NUMBER	TOT DIFF	% DIFF	ITEM NUMBER	TOT DIFF	% DIFF
101x112	19	39	112x101	30	60
102x113	18	37	113x102	32	64
103x114	26	52	114x103	31	62
104x115	22	45	115x104	30	60
105x116	16	33	116x105	35	70
106x117	19	39	117x106	33	66
107x118	19	38	118x107	38	76
108x119	17	35	119x108	40	80
109x120	18	36	120x109	37	74
110x121	24	48	121x110	35	70
111x122	21	42	122x111	41	82
112x123	27	54	123x112	29	58
113x124	20	40	124x113	43	86
114x125	24	48	125x114	34	68
115x126	22	44	126x115	32	64
116x127	19	38	127x116	27	54
117x128	23	46	128x117	24	48
118x129	24	49	129x118	15	30
119x130	17	35	130x119	19	38
120x131	10	20	131x120	12	24
121x132	18	37	132x121	18	36
122x133	11	22	133x122	12	24
123x134	6	12	134x123	10	20
124x135	5	10	135x124	5	10
125x136	7	14	136x125	10	20
126x137	5	10	137x126	8	16x