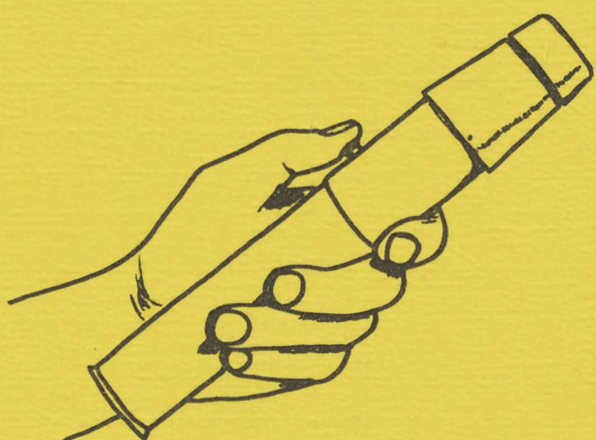
PITCH INFLECTION

IN ELECTROLARYNGEAL SPEECH



ROB C. VAN GEEL

PITCH INFLECTION IN ELECTROLARYNGEAL SPEECH

Proefschrift

ter verkrijging van de graad van doctor in de Letteren,

aan de Rijksuniversiteit te Utrecht,

op gezag van de Rector Magnificus Prof. dr. O.J. de Jong,

volgens besluit van het College van Decanen in het

openbaar te verdedigen op vrijdag 1 juli 1983

des namiddags te 2.45 uur

door

Robert Cornelis van Geel

geboren op 16 juni 1953

te Hilversum

PITCH INFLECTION IN ELECTROLARYNGEAL SPEECH

Proefschrift

ter verkrijging van de graad van doctor in de Letteren,

aan de Rijksuniversiteit te Utrecht,

op gezag van de Rector Magnificus Prof. dr. O.J. de Jong,

volgens besluit van het College van Decanen in het

openbaar te verdedigen op vrijdag 1 juli 1983

des namiddags te 2.45 uur

door

Robert Cornelis van Geel

geboren op 15 juni 1953

te Hilversum

Promotor : Prof. dr. A. Cohen

Coreferenten: Dr. M.P.R. van den Broecke

Prof. dr. P.H. Damsté

Acknowledgements

This reserach was supported by the Foundation for Linguistic Research, which is funded by the Netherlands Organization for the Advancement of Pure Research (ZWO). It was carried out at the Institute of Phonetics of Utrecht University.

I wish to thank the following persons who contributed to this dissertation in various manners:

First, my promotor Prof.dr. A. Cohen, for displaying a keen sense of perspective for main issues, which proved to be very helpful. By virtue of his 'terrier like' quality he kept track of what those concerned with this research project were doing, thereby ensuring that progress continued. I also take this opportunity to thank him for his numerous courier services and for being my fellow traveller to the IPO.

Dr. M.P.R. van den Broecke for being my supervisor and co-referent and for producing valuable criticisms on all matters concerning the present study. On top of that, he has always been available for instant and to-the-point advice.

Prof.dr. P.H. Damsté of the Institute of Phoniatrics of the Medical Faculty of Utrecht University, for being my co-referent as well as for providing valuable suggestions throughout the course of the project.

The other members of my 'workgroup', Drs. Sj. van der Meulen, Drs. J.R. de Pijper, Drs. J.C.T. Ringeling, and Dr.Ir. G.H. Wieneke, for their valuable comments, criticisms, and suggestions on various topics in this study.

Dr. V.J.J.P. van Heuven, for reading the prefinal manuscript and making many useful suggestions.

My colleagues, for many discussions on matters of phonetic or culinary interest, and, especially, Ing. C.G. van Bergh, for designing electrolarynx prototypes which, despite their highly experimental appearance, are still in working order.

A special word of thanks goes to my three speakers, of whom Mr. K.J. Haak Steenhart deserves special mentioning for his enthusiastic support with virtuoso demonstrations. The cooperation of those who served as subjects in the preliminary application of the learning program and those who served as listeners in my tests is also gratefully acknowledged.

The secretaries Mrs. A. van der Linden and Mrs. L. Taju, for assistance in typing this manuscript and being able to read my highly personal style of handwriting.

My parents, for providing support in more than one sense of the word.

Yvonne, for sharing my life and my erratic memory for everyday matters, increasingly so during my writing this dissertation.

CONTENTS

Chapter 1

DUTCH PITCH THEORY

1.1 Introduction

In many science fiction films robots and computers possess the faculty of speech. The reason why they sound so alien is almost always that their speech is monotonous. In daily life it is possible that we meet a person who produces such a type of speech (however, the comparison ends here). It will be someone who uses a so-called electronic artificial larynx - or electrolarynx for short - who has permanently lost his voice (most are males) as the consequence of an operation in which the voice box, including the vocal cords, has been entirely removed by surgery. The electrolarynx device emits a sound which serves as a substitute voice source (more on actual speech production with this device in section 2.4). The resulting speech is characterized by a monotonous quality. With today's technology it is simple to equip such a device with a control for varying the pitch of the sound emitted. In fact, this was done as early as the 1960's. Nevertheless, there have been surprisingly few people who were able to use this pitch control and really achieve natural pitch inflection. One might ask why so many people are unable to use the possibility of controlling the pitch on the electrolarynx. The present study is concerned with just that: it aims to find out why this is the case, and if a plausible reason is found, what can be done to remedy it.

In this study research carried out on Dutch intonation plays an important role. For this reason it was decided that before an outline of this study and a working hypothesis are presented, an introduction to Dutch intonation theory will be given first. In the second chapter the reader will be informed about the problems facing someone who has lost his vocal cords.

After that the aim of this study will be outlined. The description of the former is based both on the introduction to Dutch intonation theory and on the introduction to electrolaryngeal speech.

1.2 Studying pitch in accentuation

When we speak, the periodic sound source for our speech production is generated by the vocal cords. Air pressure induced by the lungs forces the vocal cords to open. A bit of air is let through and the vocal cords close again. This cycle repeats itself at a fairly high rate. The quick succession of pulses gives us the sensation of a tone. The number of times the vocal cords open and close define the fundamental frequency of the voice. The sensation of a tone is accompanied by a sensation of pitch which is the perceptual correlate of the fundamental frequency of the voice.

In this study the term pitch is not only used strictly for the perceptual domain, but also in a rather loose sense, as is common in studies on pitch phenomena, it comprises the acoust domain. For instance the pitch control of an electrolarynx is, properly speaking, a control for varying the fundamental frequency of the sound emitted.

Pitch variations of the voice play an important role in accentuation. Pitch inflection is widely recognized as the primary contributing factor in the realization of prominence (Bolinger, 1958; Fry, 1958; Denes, 1959; Morton & Jassem, 1965; Lehiste, 1970). Many researchers have attempted to try and find the intonational systems of a particular language. In this respect, the language best studied is probably English. Two main 'schools' have evolved from this research: the British and the American school.

Basically the British school as it is reflected in Crystal (1969), Halliday (1970) and O'Connor & Arnold (1973) describes the intonational system of British English with the aid of constructs called tunes or tone groups, centring round a main pitch movement 'nucleus', which may be preceded by a '(pre-)head' and followed by a 'tail'. These constructs have a close link with, on the one hand, attitudes, expressing emotions such as indignation, unbelief, happiness, etc., and, on the other hand, with syntactic characteristics of the utterance.

In the American school pitch contours are considered to be based on four levels (Wells, 1945; Pike, 1945; Trager & Smith, 1951). Again intonation is seen as linked with attitude and syntax. A discussion of the major intonational systems of both schools can be found in Willems (1982).

Many linguists have developed different notational systems. Lieberman (1965) demonstrated that the characteristics of the Trager Smith system showed inadequacies when applied to less frequently occurring sentence structures.Systems like these cannot be considered reliable, especially since they are based more often than not on impressionistic data.

In a listening experiment Brown, Curry & Kenworthy (1980) found that even trained phoneticians did not agree as to the position of the nucleus of the British English systems. It appeared that the meaning of the sentences provided the main clue for the nucleus.

It seems that the main difficulties of both the British and the American schools are caused by attempts to connect intonational systems with meaning or linguistic function and by the fact that impressionistic data were used as a basis for setting up descriptive systems which were seldom verified in experiments.

During the last twenty years in the Netherlands intonation research has been conducted which resulted in a grammar of pitch movements. In the next section I shall briefly review the basic method and assumptions underlying the Dutch school of intonation.

1.3 Perceptual analysis

The Dutch method aims at reducing the seemingly capricious pitch curves to a sequence of a limited number of pitch movements which are perceptually relevant. The notion 'perceptually relevant' will be discussed below.

1.3.1 Perceptual relevance

The human ear is highly sensitive to pitch differences. The just noticeable difference for pitch with steady state stimuli is about 0.3 Hz and for stimuli with descending ramps (105 - 140 Hz) about 2 Hz (Klatt, 1973). Nevertheless, phoneticians highly trained in listening to pitch, often have difficulties in deciding whether a pitch movement is rising or falling.

't Hart (1966) describes a method called 'analytic listening' in which consecutive, gated out portions of 30 ms of the speech utterance to be analysed were matched by ear as to pitch height with an adjustable reference signal. A period counter was used to determine the adjusted frequency. Although this method was laborious it had two advantages: lacking reliable frequency meters at the time it was a reliable method (claimed accuracy: about 1%) to measure pitch by ear. Secondly, it proved to be feasible, after some training time, for the listeners to make predictions about major pitch movements in concutive sentences to be analysed. Comparison of these predictions with the actual measurements further helped in making correct predictions.

An important feature which seemed common to all utterances was that they seemed to contain a gradual decrease in pitch over the entire utterance, which was termed 'declination'. Initially, an explanation for this phenomenon was given by Collier (1975) who found that during an utterance the subglottal airpressure decreases and that this is the main cause for the gradually declining pitch. Maeda (1976) found in his study on American English intonation that the slope of the declination line varied with the length of the utterance. Since declination is considered a by-product of exhalation, it may be assumed that, apparently, exhalation is dependent on utterance duration. Thus, deciding when to begin and when to end an utterance implicitly defines the slope of the declination line. Clearly, these decisions are under voluntary control of the speaker. In summary, it may be stated that the slope of the declination line is implicitly under control of the speaker by virtue of the decision when to begin and when to end an utterance. A discussion of these matters concerning the status of the declination phenomenon can be found in Cohen, Collier & 't Hart (1982) in which the old description by Cohen & 't Hart (1965) is revised. The status of an independent topline is also discussed. It was found to be dependent on the baseline: it runs parallel to it in a logarithmic frequency domain.

It turned out that all perceptually relevant pitch movements can be described as major deviations from the declination line. These pitch movements are supposed to be present in every sentence. Other minor pitch movements were ascribed to phoneme and phoneme sequence characteristics. It was hypothesized that the major pitch movements were perceptually relevant. This notion is coupled to the assumption that what is relevant to the listener's ear, was intended by the speaker (Cohen & 't Hart, 1965).

1.3.2 Analysis by (re-)synthesis

To find out whether the simplification obtained by describing pitch contours in terms of these major pitch movements could be improved or had been carried out already one step too far, a subsequent phase in the perceptual analysis was set up.

This was done with the aid of the Intonator (Willems, 1966). The Intonator is essentially a vocoder system which consists of a set of analysing filters dividing the frequency spectrum into adjacent bands and a similar set of filters used in synthesis. The incoming signal is analysed in the first filter set. The output of the analysis serves as input for the second set of filters where the signal is resynthesized. In the Intonator pitch information for synthesizing the fundamental frequency is not obtained through analysis but by manual input of the control parameters. Thus it was possible to resynthesize utterances with almost the original spectral qualities but with a pitch contour which could be varied at will.

The simplified contours could now be fed into the system to obtain stimuli for a listening test on distinguishing between original and simplified versions.

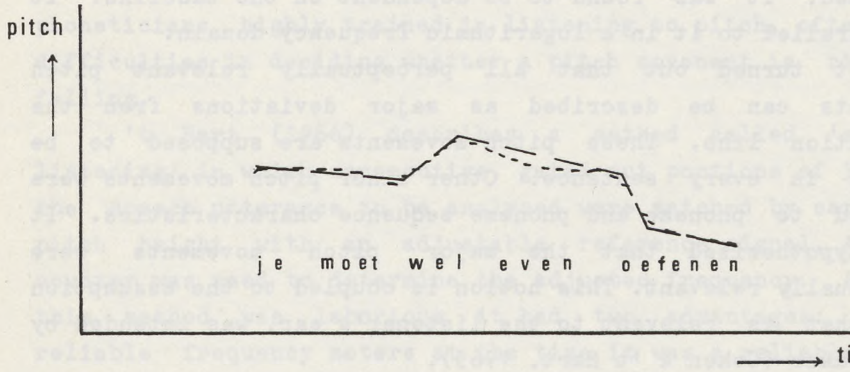


Fig. 1.1: original pitch contour (dotted)and its simplification (dashed).

Thus it was established which pitch movements were truly perceptually relevant and which were not. One observation made was that major pitch variations were placed almost invariably on those syllables in salient words of an utterance that carried lexical stress.

In later research along the same lines on British English intonation this kind of perceptually tested optimal simplification of the original pitch contour was formalized as a so-called 'close-copy stylization', which consisted of the smallest possible number of straight lines, and which is perceptually indistinguishable from the original (de Pijper, 1979, and to appear).

1.3.3 Standardization

As the perceptually relevant pitch movements were shown to be the result of voluntary articulatory gestures on the part of the speaker (Collier, 1975), it was thought that these movements could also be utilized in speech synthesis by rule. Therefore, instead of feeding the Intonator with close-copy stylizations, attempts were made to standardize the perceptually relevant pitch movements (see fig. 1.2).

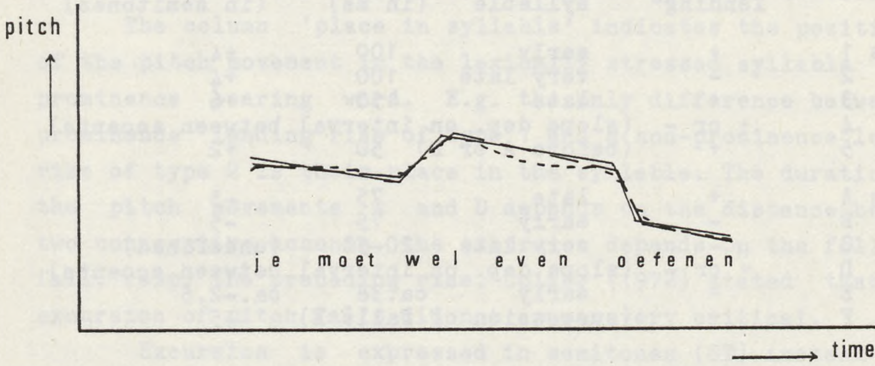


Fig. 1.2: original (dotted), close-copy (dashed) and standardized (solid lines) pitch contour (cf. fig. 1.1).

Listening tests were employed to find out whether the degree of standardization was permissible. This time, however, the question to the listeners was not whether the original contour and its standardized version were indistinguishable, but whether the standardized contours were intonationally acceptable, without direct comparison to the original.

This yielded a limited number of different perceptually relevant pith movements. The major movements in the close-copy stylizations can be seen as variants of this limited set.

Therefore, with perceptually relevant pitch movements both the standardized movements and their variants can be designated, depending on the level of description. In table I the standard values for each movement are given, excepting the declination line. These values are based on those given in 't Hart & Cohen (1973) and in 't Hart & Collier (1975).

Table I: standard values for perceptually relevant pitch movements.

	type	prominence lending	place in syllable	duration (in ms)	excursion (in semitones)
rises	1	+	early	100	+4
	2	-	very late	100	+4
	3	+	late	150	+6
	4	+ or -	(slope dep. on interval between accents)		
	5	+	(before A or D)	50	+2
falls	A	+	late	75	-5
	B	-	early	75	-5
	C	-	very late	20-50	(undefined)
	D	+ or -	(slope dep. on interval between accents)		
	E	+	early	ca.38	ca.-2.5
	F	+	(succession of falls E)		-5

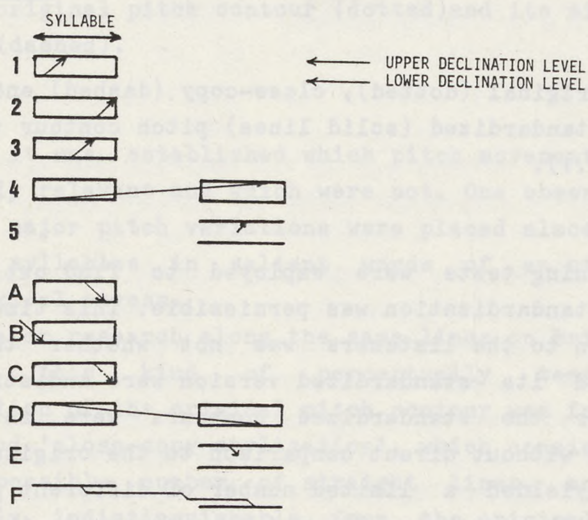


Fig. 1.3: schematized drawings of the pitch movements from table I.

For an extensive decription of most of these movements the reader is also referred to Collier (1972). Here only a cursory explanation will be given.

The column 'type' indicates the notation for each of the movements. The next column indicates whether or not a movement lends prominence to the word it appears in: it turned out that, basically, the succession of these movements was an alternation between rises and falls and that sometimes speakers appeared to produce non-prominence lending pitch movements to allow for two successive pitch accents to be realized by two rises. In the present study only rises 1, 2 and 3 and falls A, B and C were concentrated on for reasons which will be given in section 2.8.

The column 'place in syllable' indicates the positioning of the pitch movement in the lexically stressed syllable of a prominence bearing word. E.g. the only difference between a prominence lending rise of type 1 and a non-prominence lending rise of type 2 is their place in the syllable. The duration of the pitch movements 4 and D depends on the distance between two consecutive accents, the excursion depends on the following fall, resp. the preceding rise. Colier (1972) stated that the excursion of pitch falls did not appear very critical.

Excursion is expressed in semitones (ST) instead of in Hertz in order to be speaker independent, for every speaker has his or her own average overall pitch level. Furthermore, this unit has a logarithmic relation to frequency, which better reflects the way our ears react to pitch differences. Thus the distance between two tones of 100 and 200 Hz appears equal to the difference between tones of 200 and 400 Hz: both distances are 12 ST, i.e. one octave. The distance between two frequencies can be calculated as follows:

$$(\log f2 - \log f1) \times \log 2 : 12 = ST$$

The specifications of the declination line have not been dealt with so far. In early studies a decrease in pitch, expressed in Hz, of 3% for every 100 ms was proposed (Cohen & 't Hart, 1967). A more precise formula is given by 't Hart (1979):

for $t \leqslant 4.82$ s: $D = -1/(0.13 + 0.09t)$, and
for $t > 4.82$ s: $D = -8.55/t$.

Here 't' is the duration of the utterance in seconds and 'D' is the slope of the declination line expressed in ST/s. These formulae can be applied to stretches between a fall and a consecutive rise called 'low declination', and between a rise and a consecutive fall called 'high declination'. (The values for the standard pitch movements as given in table I do not include declination; this has to be added to them according to the formula chosen.)

1.4 Grammar of Dutch intonation patterns and some probabilistic considerations

As can be seen in table I some pitch movements have sequential relations to other movements, e.g. rise 5 comes directly before fall A or D. It has already been observed (Cohen & 't Hart, 1967) that the succession of a fast rise and a fast fall (1A) together with stretches of low and high declination appears very often. This is called the 'hat pattern'. Many other patterns seem to be derived from this basic pattern. Apparently there are particular sequential relations between pitch movements, possibly subjected to some hidden grammar.

In models set up to find out whether there are rules which order the successions of pitch movements (Collier, 1972) it turned out that they can be taken together into intonational blocks: certain successions of pitch movements appeared to be strongly coherent. Three types of blocks are distinguished: end blocks (E) which are always present in a pitch contour, with or without other blocks; continuation blocks (C) which occur before major syntactic breaks; and prefix blocks (P) which can precede both other types. The relative frequency of occurrence of these blocks was determined on the basis of a corpus consisting of one and a half hour of spontaneous speech and six samples of about ten minutes of very colloquial and fluent stage speech. Figure 1.4 depicts the transitional probabilities between the three types of blocks (taken from Collier, 1972).

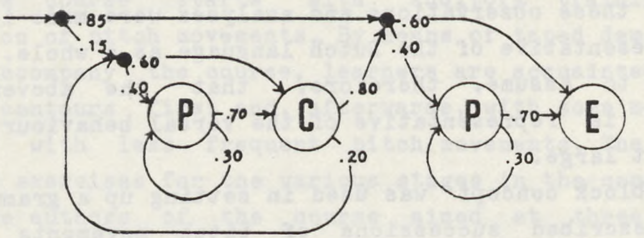


Fig. 1.4: transitional probabilities between prefix (P), continuation (C), and end (E) blocks.

It will be apparent that pitch patterns consisting of few blocks are far more frequent than patterns with more blocks. Colier (1972) also gives the relative frequency of occurrence of the various pitch movements for each type of intonational block. Based on these data the relative frequency of occurrence of pitch patterns consisting of pitch movements with the notations 1, 2, 3, A, B, and C is given in table II for those combinations which occur at least in 1% of all cases.

Table II: relative frequency of occurrence of pitch movement combinations.

sequence	percentage
1A	33.2
1B1A	14.6
3C	6.4
1B1B1A	4.2
	3.9
1B3C	2.8
A	1.9
1B1	1.7
1B1B1B1A	1.2
1A2B1A	1.0

This set of pitch movement sequences, which can be described as the hat pattern and its derivations, appeared in about 70% of all sentences which had been analysed. The corpus on which these observations and analyses were made is claimed to be representative of the Dutch language as a whole. It seems plausible to assume, therefore, that the abovementioned percentage is representative of the verbal behaviour of Dutch speakers at large.

The block concept was used in setting up a grammar. This grammar described successions of pitch movements as they combine into blocks and it accounted for more than 95% of the contours observed in the speech samples analysed.

1.5 Can the intonation grammar be taught for practical purposes?

Point of departure in the Dutch school of intonation research is the assumption that those pitch movements which are perceptually relevant are in fact intended by the speaker. A description which comprises all these movements can therefore be considered to have a psychological reality ('t Hart, 1981).

When Dutch is learnt as a second language, its intonation may possibly be mastered in a cognitive learning process, rather than by the usual listen-and-repeat exercises. Such a learning process depends on study and analysis of the subject to be learnt and on conscious control over what is learnt. The explicit description of Dutch intonation as laid down in the intonation grammar appears to be suitable for application in a cognitive learning process.

Before starting to learn the underlying system of intonation, however, the second language learner of Dutch should receive training in listening to pitch and its variations. The learner should become conscious of which types of patterns are possible in Dutch. This perceptual 'pitch consciousness' is used as the basis for the production of pitch contours.

These considerations together with the intonational grammar were incorporated in the development of a 'Cursus Nederlandse Intonatie' (course in Dutch intonation) by Collier & 't Hart in 1978 (published in 1981).

The course starts with auditory training in the perception of pitch movements. By means of taped demonstrations which accompany the course, learners are acquainted with some simple contours first and, afterwards, with some more complex contours with less frequent pitch movements. The book also contains exercises for the various stages in the course.

The authors of the course aimed at three groups of potential users: learners of Dutch as a foreign language, students of Dutch linguistics, and possibly perceptually handicapped native speakers of Dutch ('t Hart & Collier, 1978). Although to this date the third application has not yet been effected, it was thought that also persons with a production handicap could benefit: users of an electrolarynx who can only produce monotonous speech. This type of speech often elicits comments as: robot speech, it sounds like a talking computer, etc. It was thought that speakers with this productional handicap, even when given the technical facilities, still did not know how to vary pitch. They have an advantage over second language learners: Dutch intonation is not new to them, they are simply not conscious of the pitch variations they were able to produce before their operation. A cognitive approach might conceivably remedy this.

In the second chapter a survey is given of the speech rehabilitation possibilities for people whose larynx has been excised. In it an account is given of some aspects of electrolaryngeal speech production and its monotony. In the third chapter an intonation learning program together with its application is described, which is strongly based on the principles of the intonation course mentioned above. The next chapter deals with a perception based evaluation of the quality of intonated electrolaryngeal speech. In chapter five conclusions from this study will be drawn and suggestions for further research will be made.

This set of pitch movement sequences, which can be described as the hat pattern and its derivations, appeared in about 70% of all sentences which had been analysed. The corpus on which these observations and analyses were made is claimed to be representative of the Dutch language as a whole. It seems plausible to assume, therefore, that the abovementioned percentage is representative of the verbal behaviour of Dutch speakers at large.

The block concept was used in setting up a grammar. This grammar described successions of pitch movements as they occurred in blocks and it accounted for more than 99% of the contours observed in the speech samples analysed.

Can the intonation grammar be taught for practical purposes?

Point of departure in the Dutch school of intonation research is the assumption that those pitch movements which are perceptually relevant are in fact intended by the speaker. A description which comprises all these movements can, therefore, be considered to have a psychological reality ('t Hart, 1981).

When Dutch is learnt as a second language, its intonation may possibly be mastered in a cognitive learning process, rather than by the usual listen-and-repeat exercises. Such a learning process depends on study and analysis of the subject to be learnt and on conscious control over what is learnt. The explicit description of Dutch intonation as laid down in the intonation grammar appears to be suitable for application in a cognitive learning process.

Before starting to learn the underlying system of intonation, however, the second language learner of Dutch should receive training in listening to pitch and its variations: the learner should become conscious of which types of patterns are possible in Dutch. This perceptual 'pitch consciousness' is used as the basis for the production of pitch contours.

These considerations together with the intonational

Chapter 2

SPEECH AFTER LARYNGECTOMY

2.1 Introduction

This second introductory chapter will deal with the nature of the handicapped population of laryngectomees, their immediate vocal problems, a brief description of solutions for these problems including recent state-of-the-art rehabilitation possibilities. Special emphasis will be laid on electrolarynges and on one of their main limitations: the monotony in speech produced with these devices. After hypothesizing about the reason for this monotony an outline will be presented of a learning program to overcome this monotony, and of the development of a prototype EL with semi-automatic pitch control to facilitate relatively easy control of the device.

2.2 Description of the population

In the majority of cases the medical history of laryngectomees, patients who have had their larynx taken out, amounts to the following: most patients visited their doctors with complaints like a sore throat, hoarseness, swallowing difficulties, which would not pass over. The patients were sunsequently referred to a speech laryngologist who had a biopsy performed on tissue of the vocal cords. After the presence of a carcinomic tumor in the larynx had been established, it was decided to perform a total laryngectomy, an excision of the complete voice box. This was often combined with pre-operative and post-operative X-ray treatments to preclude further development of suspect cancerous tissue in the immediate vicinity of the larynx.

Each year about 200 laryngectomies are performed in the Netherlands (Damsté, personal communication). Recent figures on laryngectomies in America amount to 10.000, annually (Sekey,1982). With an increasing average age these numbers are likely to increase.

Conclusive evidence exists that heavy smoking is conducive to cancer of the larynx. This opinion is expressed by Snidecor (1962, p.14): '...if the case against the safety of the Brooklyn Bridge was as tight as the case against the safety of smoking, it would be closed to traffic at once'.

The result of a laryngectomy is not limited to the loss of the larynx and its vocal cords, the trachea is bent forwards and sutured to an artificially made opening, a stoma, in the neck a litle above the sternum. Fig. 2.1 illustrates the situation before and after laryngectomy.

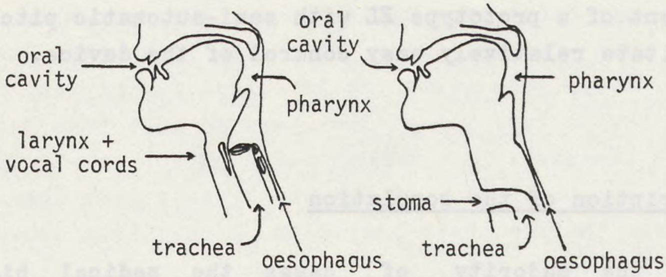


Fig. 2.1: midsagittal section of head and neck before (left) and after (right) total laryngectomy.

Breathing is done through this stoma and there is no connection between the oral cavity and the lungs. This, and the loss of the vocal cords, deprives the postoperative patient of the main source for speech production: the voice.

Directly after the operation the patient's only way of communication is writing. Often a 'magical slate' is used for this purpose. This often slows down communication to the point where the visitors are not patient enough to wait for the patient to finish writing his message.

2.3 A survey of speech rehabilitation possibilities.

2.3.1 Introduction

Basically three ways of speech rehabilitation are open to the patient, depending on anatomical possibilities, on the patient's personal preferences and skill, and/or motivation to spend a lot of effort to master a certain speech mode. These ways are: 1) intrinsic forms of alaryngeal speech, and 2) extrinsic forms of alaryngeal speech in the form of surgical-prosthetic methods of speech restoration, or 3) in the form of artificial larynges (terms taken from Weinberg, 1980). Of these forms two specific modes of speech are most fequently used in alaryngeal speech rehabilitation: oesophageal speech and electronic artificial larynx speech, in that order. However, a short account will be given of most types of speech rehabilitation, with the main emphasis on artificial larynges. For an extensive historical survey of the development of rehabilitation possibilities for the laryngectomized see Lowry (1981).

2.3.2 Intrinsic forms of alaryngeal speech

Intrinsic forms of alaryngeal speech are those forms which can be realized by the intrinsic anatomical structures which remain after laryngectomy. They are: buccal, pharyngeal, pseudo-whispered, and oesophageal speech, of which the latter is the most widely known and applied form of speech rehabilitation for laryngectomees.

Application of buccal speech is rare (Diedrich & Youngstrom, 1966) and is indeed considered undesirable and unpractical as a primary mode of alaryngeal speech (Damsté, 1958; Lauder, 1968). In buccal speech the air supply comes from within the oral cavity with cheek and jaw forming a neoglottis (Weinberg & Westerhouse, 1971). This sort of speech can only be understood by the initiated.

Pharyngeal speech is produced with the tongue serving both as an articulatory and vibratory organ, the air supply being supplied from the pharyngeal cavity. A detailed study of one pharyngeal speaker, dealing with the physiological speech mechanisms and speech intelligibility, is reported in Weinberg & Westerhouse (1973). The desirability of this type of speech is rated at the same low level as buccal speech.

Pseudowhispered speech is produced with buccal air while the bottom of the mouth is used as a pump (Damsté, 1975). Sometimes buccal speech is also referred to by the term pseudo-whisper (Malbeck & Schosshauer, 1960). However, buccal speech is characterized by a croaking voice (Damsté, 1975), which is unlike whispered speech; it seems desirable, therefore, to use different terms for both speech modes.

Oesophageal speech is considered the best intrinsic form of alaryngeal speech and indeed it is the type of speech which is almost always tried first in the rehabilitation of the laryngectomee.

To produce sound in oesophageal speech air must first be passed into the oesophagus. Different methods to achieve this have been described.

When the injection method is used, the laryngectomee uses movements of the floor of the mouth to increase intra-oral and pharyngeal pressure which causes the pharyngoesophageal segment to open and to allow air in the oesophagus. Air injection can be accomplished in two ways. One way is the so-called consonant injection method in which use is made of pressure buildup in the production of voiceless plosive consonants, especially /p/ and /t/ (Moolenaar-Bijl, 1953). A second way of injection is the glossal or glossopharyngeal press. This method 'consists of pumping air into the lungs through coordinated stroking actions of the tongue, jaw, and pharynx' (Weinberg & Bosma, 1970).

In the inhalation method (described as early as 1925 by Burger & Kaiser) '...the patient inhales pulmonary air, there is a subsequent decrease in the negative presssure that normally exists in the esophagus....Atmospheric air pressure in the mouth and hypopharynx will push air into the esophagus...' (Diedrich, 1968).

Once air has passed into the oesophagus through one of these methods, it has to be expelled to achieve phonation. Van den Berg, Moolenaar-Bijl & Damsté (1958) describe two main mechanisms for air expulsion. 1) A force coming from the elasticity of the walls of the oesophagus which helps in maintaining a constant pressure for a steady flow of speech. 2) The intra-tracheal pressure rise due to exhalation which is transmitted through its walls to the oesophagus. This pressure is enough to produce a louder voice or a longer phrase. The pharyngoesophageal sphincter is opened by the sub(neo-)glottal pressure and set into vibration. This vibration then serves as a voice source.

2.3.3 Extrinsic forms of alaryngeal speech: surgical methods

Extrinsic forms of alaryngeal speech are those forms which rely on surgically made structures created for voice production. All methods are characterized by a surgical connection between the trachea, which was closed in the laryngectomy operation (cf. fig.2.1), and the vocal tract. Some methods provide an air shunt between trachea and oesophagus or pharynx, other methods provide both an air connection and a mechanical voice source between trachea and pharynx. Most of these methods never had any wide application, with possible exceptions of the Asai technique and the Staffieri technique, although it is hard to say which one of these methods has become truly successful.

The Asai technique has been developed in Japan and is described by Miller (1962). This operation provides an airway in the form of a dermal tube between the trachea and the lower pharynx. Upon closing the stoma with the finger the patient could the breathe air via this tube into the pharyngeal cavity. Asai speech is learned much faster than oesophageal speech (Curry, Snidecor & Isshiki, 1973). The source of voicing is not described. This method requires three extra operations, which form an additional burden to the patient.

Another technique makes use of the 'LaBarge VoiceBak' prosthesis. A second stoma in the neck, connected to the oesophagus is needed for this technique which is the result of research conducted by Taub (1975) and associates. The prosthesis provides an external air bypass from the tracheal stoma to the oesophageal stoma. The voice source is the same as in oesophageal speech, therefore it is not really a voice prosthesis, as the name of the device suggests. A similar device is the North Western voice prosthesis, as developed by Sisson, Mc Connel, Logemann & Yeh (1975), which, again, is but an air bypass. However, this device is connected to a hypopharyngeal fistula.

The method developed by Staffieri differs from the above mentioned methods in that no air tube of any kind is utilized. Instead a fistula is created through the wall between trachea and oesophagus. A brief description is given by Vega (1975) and Blom & Singer (1979). As in the other methods, air reaches the oesophagus below the pharyngoesophageal segment where the vibration can be induced which serves as a voice source.

The Staffieri technique proved to have limited usefulness in the irradiated patient, caused chronic swallowing difficulties and a success rate which approximated that for oesophageal speech (viz. 68%). This led Blom & Singer (1979) to develop a method based on a tracheoesophageal puncture combined with a valved tube. This allows air to pass into the oesophagus and it keeps fluids from entering the trachea. The device is known as the duckbill speech tube or the Blom-Singer voice prosthesis. This tube is made of silicone and has two retention flanges at the front end which are taped to the skin next to the stoma. The tube is open both at the front end and on the side. The other end of the tube is round and has a razor-thin slit which serves as a one-way valve (see also fig. 2.2).

The tube is inserted into the stoma and directed into the tracheoesophageal puncture behind the stoma, so that the valved end is in the oesophagus. The retention flanges help to keep the duckbill tube in its place. When the speaker exhales while closing his stoma with a finger, the air enters the tube through the airflow port and reaches the oesophagus. Again, as stated in the above mentioned methods, an oesophageal voice is created.

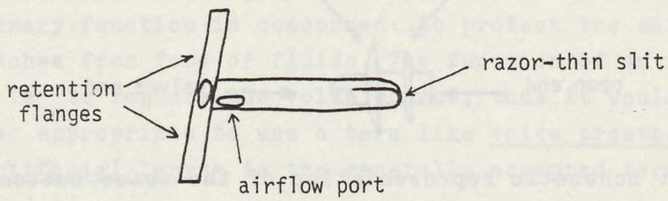


Fig.2.2: schematic illustration of the duckbill speech tube.

All methods which direct pulmonary air into the oesophagus yield an oesophageal type of voice with one big difference: the air supply can be controlled much better to the extent that more syllables can be pronounced at a stretch and that a better voice quality can be obtained with respect to modulation of pitch and loudness, for the same source of air pressure, viz. the elasticty of the lungs, is available to produce exhalation pressure as was the case with the natural voice. However, due to the very high resistance of the valve the pressure in the trachea is much larger than in the oesophagus. Yet, it stands to reason that this provides a better control over the air pressure under the pharyngo-esophageal segment than the ways of air expulsion described in section 2.3.2.

Singer, Blom & Hamaker (1981) report a 88% success rate with 129 patients, but others report 56% with 23 patients. (Donegan, Gluckman & Singh, 1981).

Some patients had difficulties in keeping the duckbill valve in its place. This caused Panje (1981) to develop the 'voice button' which serves the same purpose, but has two flanges around the tube which is to be positioned such that both flanges are on opposite sides of the tracheoesophageal wall (cf. fig. 2.3).

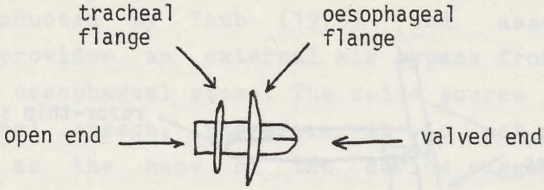


Fig. 2.3 : schematic representation of the voice button.

The open end is in the trachea, not in the stoma. This device operates along the same principles as the duckbill device.

Independently, an air valve very similar to the voice button has been developed in Groningen, the Netherlands, by Nijdam, Annyas, Schutte & Leever (1982); it is called the 'Grunneger sproakknuppie' (Groningen voice button).

Two modifications of the Blom-Singer duckbill valve have been made by Shapiro & Ramanathan (1982), and by Perry, Cheesman & Eden (1982) (the 'Charing Cross modification'), which both direct pulmonary air into the pharynx (the reasons for these modifications were different). In neither case the site of the neoglottis, nor the quality of the voice is discussed.

Generally speaking, it can be said that the tracheo-esophageal puncture method is a promising alternative for oesophageal speech. It can also be applied in those cases where a weak oesophageal voice is present, or when only a few syllables can be pronounced at a stretch.

Panje, VandeMark & McCabe (1981) mention that they have a success rate of 80% for a group of unselected patients. This figure may be considered as indicative of the expectation that this method, however promising, is not suitable for every laryngectomee. This still leaves room for a speech mode which is at present the second most frequently used way of speech restoration: the electronic artificial larynx.

2.3.4 Extrinsic forms of alaryngeal speech: artificial larynges

Most artificial larynges are not what the name suggests. They do not replace the natural larynx, at least not as far as its primary function is concerned: to protect the entrance of the trachea from food or fluids. The function of an artificial larynx is to replace the voice source; thus it would probably be more appropriate to use a term like <u>voice prosthesis</u>, but since <u>artificial larynx</u> is the generally accepted term, it will be adhered to.

The first successful laryngectomy was performed by Billroth in Vienna in 1873. He had an internal pneumatic artificial larynx made for his patient. A pneumatic artificial larynx is powered by pulmonary air which sets a reed into vibration. This internal artificial larynx is positioned in the place of the natural larynx with the curved end in the trachea and the tube containing the reed directed towards the pharyngeal cavity (cf. fig. 2.4).

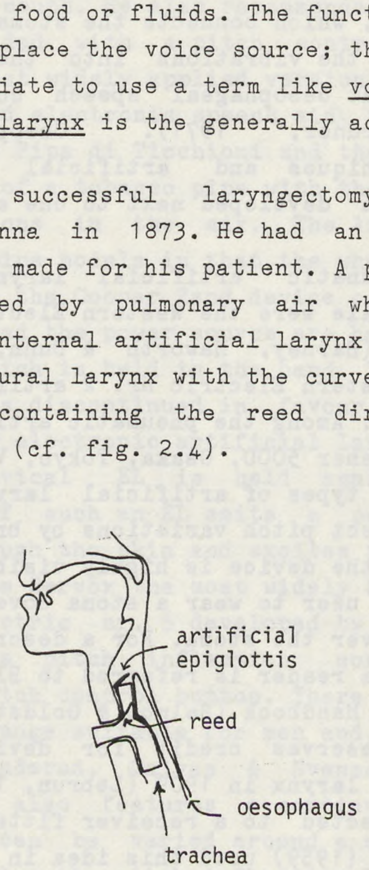


Fig. 2.4: internal pneumatic artificial larynx in situ (schematized).

The above figure also shows an artificial epiglottis; therefore this device could be considered a true artificial larynx. Whether this artificial epiglottis really functioned satisfactorily, i.e. closed upon swallowing food or liquid, is debatable (Lebrun 1973). Some other physicians followed this example in essence. Around the turn of the century the first external pneumatic larynx was devised by Gluck. Devices of this kind usually consist of some sort of a flexible tube, with a vibrating reed in it, which connects the stoma with the mouth. A small pipe delivers the vibrations into the mouth. In 1908 Gutzmann had shown that oesophageal speech could be taught to laryngectomees (Gardner, 1971). Oesophageal speech rehabilitation techniques and artificial larynx speech techniques have been developed next to one another from then on.

The first pneumatic artificial larynges which were applied on a large scale were the Western Electric nr. 1 (1925) and nr. 2 (1929) (Barney, Haworth & Dunn, 1959). A full description of the Western Electric nr. 2 artificial larynx is given by Riesz (1929). Among the pneumatic artificial larynges currently sold are: Neher 5000, Osaka, Tokyo, Van Hunen DSP8. An advantage of these types of artificial larynges is that it is possible to effect pitch variations by breath control. A disadvantage is that the device is highly visible and that it is impossible for the user to wear a stoma cover, as the device has to be placed over the stoma. For a description of these arificial larynges the reader is referred to Blom's chapter in the Artificial Larynx Handbook (Salmon & Goldstein, 1978).

Gluck also deserves credit for devising the first electronic artificial larynx in 1909 (Lebrun, 1973). An Edison phonograph was connected to a receiver fitted in a patient's denture. Tait & Tait (1959) used this idea in a device which, instead of a phonograph, utilized a transistorized oscillator built in a denture together with a vibrator. Recently this idea of an intra-oral electrolarynx in a denture has been taken up again by Katz (1982), Sekey & Hanson (1982), and Zwitman & Knorr (1982).

A special kind of intra-oral electrolarynx is the implanted electrolaryngeal prothesis developed by Everett & Bailey (1982). This device consists of an implanted transducer (4.3 cm diameter) together with a power-pickup coil and an external electronic package. Power transfer is realized through induction. At the moment of presentation of the paper referred to implantation had only been performed on test dogs.

Firestone (1940) designed an electrolarynx (henceforth: EL) which delivered the pulses produced by the device via a tube into the mouth. He also recommended that the oscillator should be provided with a pitch control for inflectional purposes. The most widely applied version of Firestone's idea is the Cooper Rand electronic speech aid. Some other devices of this kind are the Pipa di Ticchioni and the Danapipe, which are both in the form of a tobacco pipe with the pipe stem as a duct for the vibrations in the air. The Artivox speech aid is similar to both pipe models in that the whole device is held in the hand, whereas the Cooper Rand device consists of two parts: the electronics and the power source are housed separately from the transducer which is held in the hand. The production of the Artivox device was discontinued in favour of a transcervical model: the Servox electronic artificial larynx.

A transcervical EL is held against the neck. The vibrating end of such an EL emits a pulse which is then transmitted through the skin and excites the air in the speech tract. Next to the Servox the most widely known neck-type EL is the Western Electric nr. 5 developed by Barney et al.(1959), which features a pitch inflection control operated by depressing the pitch control button. There are two models: nr. 5A with a pitch range suitable for men and nr. 5B for women.

In 1974 Branderud, Galyas & Svensson reported on an EL prototype which also features a continuous pitch inflection control. Pitch can be varied around a neutral frequency by pushing or pulling the pitch inflection loop for lowering or raising pitch. The device also features a continuous volume control meant to be used during speech. However, in practice this control was used as a kind of sophisticated on/off switch (Galyas, personal communication). The device consists of three parts: the electronics package, carried on the body, the pitch inflection and volume control control, which are handheld, and the vibrator fitted on a neckband. Initially a vibrator by Sama (for which see below) was utilized, but recently a rubber encased vibrator was developed at KTH, Stockholm, with less direct radiation when compared to the Sama vibrator (Galyas, idem). Galyas, Branderud & McAllister (1982) state that in 6 and 9 half hour training sessions two EL speakers were able to use their EL with some success when asked to differentiate between words with different pitch contours. This device is not commercially available.

Some other commercially available EL's which are less widely known, compared to the Western Electric and the Servox, are the Aurex Neovox, the Kett Mark III, the Bart's vibrator, the Sama Electronvox, - Mini Electronvox, - Vibrovox, and - Phonatron. The latter device consists of two parts: a power source with electronics and a vibrating head which is worn on the neck with the aid of a neckband. For surveys and illustrations of most of these devices the reader is referred to Lebrun (1973) and Salmon & Goldstein (1978).

The Aurex Neovox can be fitted with an appliance consisting of an elastic cup, fitted over the vibrating head of the EL, and a flexible airpipe to allow intra-oral use of the EL. For the Servox EL a similar construction is marketed in the Netherlands by the Servox agent. The advantage of such an intra-oral conversion kit is that the newly laryngectomized patient can speak immediately after the operation and continue to use the same device when the healing of the neck tissues allows this. It is obvious that this greatly enhances communication for the patient at a moment when it is most needed (Damsté & Hosang-Jacobson, 1978).

2.4 Electrolaryngeal speech production

Artificial larynx speech is produced, in all cases, neck-type or intra-oral, by pantomiming speech production. In the case of vowels this hardly poses any problems. Consonants, however, require some adaptations in their production. Regardless whether consonants are voiced or unvoiced, most of them are produced with noise generated at certain places in the speech tract. In natural speech pulmonary air is the source for this noise, but since in the laryngectomee no pulmonary air reaches the oral cavity, a different source for this noise has to be found. The air needed in the production of consonants is obtained by compressing air which is already in the oral and pharyngeal cavities, behind the partial or complete closure which is at the place of articulation of the fricative, resp. plosive consonant. This can be achieved by moving the floor of the mouth upwards, a mechanism similar to that used in the production of pseudo-whispered speech (cf. section 2.3.2).

Differentiation between voiced and unvoiced consonants is often hard to realize, since the EL is kept switched on during the whole utterance. Isshiki & Tanabe (1972) suggest that higher than normal intra-oral airpressure is required for producing voiceless consonants 'loud enough not to be masked by the continual electrolarynx sound.' (cf. section 3.1.6).

Gardner (1971) and Rothman (1978b) advocate training in differentiating between voiceless and voiced consonants in word initial and word final positions by appropiately switching the EL on or off a little later in the case of voiceless consonants, thus creating artificially made voice onset, or offset lags. This training seems hardly worthwhile for three reasons. Word recognition errors due to consonant confusions are not very likely, because coarticulation cues can very often rule out one of the two cognates. Differences in intra-oral airpressure may cause the 'unvoiced' consonants to have a noise component strong enough to mask the EL 'voice'. Thirdly, it is obvious that there are far more consonants which are intervocalic than in an initial or final position, especially in normal speech which is in fact <u>connected</u> speech. This excludes a possibility like switching the vibrator on a little later, except in the relatively few positions directly after silent pauses between utterances.

Weiss, Yeni-Komshian & Heinz (1979) found that most consonant confusions obtained in their study were attributable to voicing errors on voiceless stops. They found rather high confusion percentages in these cases. This can be explained by the fact that their material consisted of <u>isolated</u> words in a 'rhyme test': words differing in, among other things, the presence of voice in one of the constituting phonemes. It is very rare in normal speech for two words to occur shortly after each other which only differ in voicing of one consonant, e.g. dip vs. tip, as is the case in such a test. Furthermore, presenting isolated words without any context cannot be considered equivalent to normal speech. Therefore, the high confusion percentages in this experiment need not be taken as representative of EL speech.

As could be expected, vowel confusions occurred very scarcely in the experiment by Weiss et al.(1979). The same was observed in an experiment by Van Geel (1980) in which vowels were compared that had been produced with the aid of a Servox and with the aid of a Western Electric nr.5A. The latter were misperceived slightly more frequently, possibly due to a pronounced intensity peak around 800 Hz in the emitted spectrum of the Western Electric vibrator (cf. Smith, 1962), which could have influenced the resulting vowel spectrum.

Generally speaking it can be stated that EL speech need not be hard to understand. It seems that intelligibility is highly dependent on the articulatory precision of the speaker.

Although speaking with an EL can be learnt very quickly, Snidecor considers it advisable to provide speech therapy aimed not only at having the EL speaker master the abovementioned production mechanisms which are different from natural speech, but also at helping the patient in locating and maintaining the optimum spot for placing the EL against the neck to have minimal damping of the emitted pulse by the transmitting tissues. Another point to be stressed is that phrasing must be learnt, i.e. the EL must be switched off in pauses to avoid that the EL speaker keeps 'rattling on'. In fact Gardner (1971) provides a detailed outline for training the laryngectomee in the use of the EL. On the other hand repeatedly switching the device on and off, e.g. after each sylable as is advised in the instruction leaflet of the Servox EL, truly renders 'robotlike speech', which is one of the complaints about EL speech. The speech therapist should teach the EL speaker to avoid both extremes and to phrase his sentences like in natural speech.

2.5 Electrolaryngeal and oesophageal speech

This is not the place to go into the controversy about the relative merits of electrolaryngeal vs. oesophageal speech. The reader is referred to Lauder (1968) who gives a compilation of the views for and against EL speech. Nowadays EL speech is becoming more and more accepted as a speech mode which can serve equally well as oesophageal speech (Damsté, 1975; Salmon & Goldstein, 1978).

Damsté mentions a success rate of 80% in the Netherlands for oesophageal speech, but this percentage includes speakers with poor intelligibility, poor quality of voice, or reduced intensity. Recent papers by Gates & Hearne (1982) and by Schaefer & Johns (1982) mention very low success rates of 26% resp. 24% with the laryngectomees they observed (in Texas). However, these percentages cannot be compared since the quality and intensity of rehabilitation programs differ from country to country. It can be said that in the Netherlands the success rate based on the acquisition of functional oesophageal speech is about 75% (Damsté, personal communication).

Comparisons between EL speech and oesophageal speech show divided opinions about preference for either type of speech where intelligiblity and acceptability are concerned. For a survey of investigations dealing with these comparisons see Cohen, van den Broecke & van Geel (to appear). One of the main influencing factors for a choice either way was the degree of familiarity with the various types of speech: speech pathologists preferred oesophageal speech, naive listeners preferred EL speech (McCroskey & Mulligan, 1963; Shames, Font & Matthews, 1963; Green & Hults, 1982).

2.6 Limitation of electrolaryngeal speech: monotony

Over the years many the same objections have been raised against EL speech. One of the most prominent was its monotony (Barney, Haworth & Dunn, 1959). But after the Western Electric nr. 5, which was the only commercially available EL which featured a pitch inflection control, had been marketed, monotony still was a characteristic of EL speech (Snidecor, 1962; Van Riper, 1972; Rothman, 1978a). The fact that the pitch control of the Western Electric nr. 5 (A and B) was hardly ever used led the manufacturer to replace that model by one (nr.5 C) which is only equipped with an on/off switch and an internal preset pitch control (IAL News, 1983, vol.1).

Bennett & Weinberg (1973) also state that monotony 'is often present in less proficient alaryngeal speakers'. In their experiment they use 'superior artificial larynx speakers' who 'were able to effect perceptually identifiable variations in pitch...' This is one of the few instances where apparently successful use of the pitch inflection control is mentioned, although no strict criterion was used for the selection of these speakers other than that 'they exhibited clearly superior speech...'

Gardner (1971) even gives some advice on how to use the inflection control, but he does not state when and how to vary pitch other than to depress the pitch control button to emphasize a word. This is no more enlightening than the instruction leaflet accompanying the former Western Electric EL, which stated that in order to emphasize a word, pitch must be raised momentarily. As in the twenty years of the existence of the Western Electric EL the majority of its users still produce monotonous speech, it is obvious that these instructions were of no avail. To my knowledge, and apart from the few instructions in Gardner's book, no systematic training for pitch inflection has been published to this date.

An exception to this statement could be Rothman (1978b) and Salmon (1978) who advocate inducing pitch variations on EL devices lacking an inflection control by varying the coupling pressure to the neck tissues. This effect had been observed by Rothman (1978a). On the demonstration tape accompanying the Artificial Larynx Handbook (Salmon & Goldstein, 1978) an example is presented of EL speech being varied in pitch by changing coupling pressure. It would seem however, that the speech pathologists' ears have been deceived in that a change in timbre was taken for a change in pitch: although a perceptible change in 'voice' quality was achieved, pitch variation was limited. The application of this method will probably remain reserved for a small number of extraordinarily skilled EL speakers. Especially when seen in the light of the age of the majority of laryngectomees and their possibly poor manual dexterity this number is likely to be small. It may be concluded therefore that this method is not adequate to achieve pitch inflection.

In their study on the perception of intonational contrasts in alaryngeal speech Gandour & Weinberg (1982) found that these contrasts could be achieved by oesophageal and tracheoesophageal (Blom-Singer) speakers and one user of a Western Electric EL. The Servox speaker and the other Western Electric speakers they observed were generally unable to achieve intonational contrasts. Apparently only one Western Electric speaker knew what to do with the pitch inflection control on this device, the others did not, although they had been selected as proficient EL speakers.

The analysis of EL speech and its characteristics as regards speech proficiency by Rothman (1978a) yielded the counterintuitive result that frequency range was <u>not</u> an important parameter for distinguishing between good and bad speakers. It seems plausible to assume that the Western Electric users, be they good or bad speakers, did not know what to do with the inflection control.

The next logical step would be to assume that training in the use of the pitch inflection control is the way to overcome the problem of how to produce suitable pitch variations. However, before dealing with the productional aspect of pitch inflection the perceptual aspect should be taken care of first: in order to be able to execute instructions to produce specific pitch movements, a trainee ought to be able to hear pitch movements, to be aware of pitch and know how it can vary in a perceptually relevant manner.

The present study was set up to find out wether these assumptions could indeed lead to the desired result: properly intonated EL speech. The hypothesis for this study now can be phrased as follows:

Conscious knowledge of the perceptual and productional aspects of pitch inflection is necessary for adequate use of a pitch inflection control on an EL.

2.7 Aim and limitations of the study

The aim of the study was to investigate the feasibility of instructing and training a number of EL speakers to master pitch inflection. Three factors mentioned in the hypothesis were concentrated on:

a. consciousness of perceptual aspects of pitch
b. acquisition of some knowledge of Dutch intonation theory
c. consciousness of productional aspects of pitch.

It was decided to utilize the 'Cursus Nederlandse Intonatie' (Dutch Intonation Course) by Collier & 't Hart (1981), since it deals with exactly those factors. The course starts with training in perceiving pitch movements typical of Dutch intonation. Then it provides some information on some combinations of pitch movements resulting in highly frequent pitch patterns. Furthermore it includes exercises in producing these pitch patterns. The authors state that the course has a twofold aim: 1) to have the user of the course learn to hear the relevant characteristics of Dutch intonation which helps 2) to define a perceptual goal that serves as a reference in the production of intonation. The same principle applies to the training program which was set up in the present study, even more so while the user of the EL has to be more conscious of his intonation than before the laryngectomy operation when speech production and therefore also the production of intonation came about almost 'automatically'. The training program was based on the intonation course, but was directed towards the needs and production possibilities of the EL speakers. In practice this meant that only part of the intonation course was used (cf. section 3.2).

As a training program for pitch inflection control had not been attempted before elsewhere, this study has the character of a pilot study. Therefore it seemed justified to use only a few EL speakers as subjects in this study to try and verify the hypothesis. There were practical reasons for the limitation of the number of EL speakers. As it was not clear how much time it would take to train them, this could have implied that they had to come very often to attend training sessions. Therefore it was decided to use only subjects who lived in the close vicinity to limit transportation problems.

Another limitation was made in the number of different pitch movements included in the training program. Considering the fact that the laryngectomized population consists mostly of elderly people who cannot be expected to have the manual dexterity required for producing the subtle differences between many of the different pitch pitch movements present in Dutch intonation, it seemed necessary to limit the number of pitch movements to the most frequently occurring movements only. This consideration and the wish to have an EL device which is simple to control in order to make speech production again as 'automatic' as possible, led to the development of an EL prototype with semi-automatic pitch control.

2.8 Semi-automatic EL prototype

The wish for an EL more simple to control than the Western Electric nr.5 was the consequence of the observation that over the twenty years of its existence most users did not learn to use its pitch inflection control functionally. In our opinion, the reason for this failure consisted not only in that they were never taught how to use it, but also in the very complex movements of the controlling finger required in the production of pitch movements. To see what degree of complexity is required, consider fig. 2.5.

Suppose a movement as depicted in this figure is made with a Western Electric EL. This device features a pitch inflection control which has to be depressed to raise the pitch; the control is continuous. In itself this type of control is very straightforward, but it allows many unnatural pitch movements to be made. The user has to take care that a

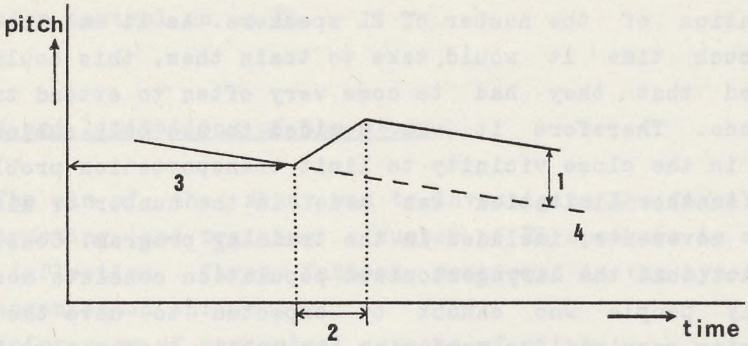


Fig. 2.5: four factors present in the production of a single pitch movement.

pitch movement is not unnaturally large or small in excursion (1), that it does not take too much or too little time (2). One has to be aware when to make a pitch movement (3), i.e. which syllable of which word should bear the accentuation induced by that pitch movement. If declination (4) is also to be incorporated the EL speaker faces an almost impossibly complex task. The user has to depress the pitch inflection control to a certain extent and release it gradually while speaking; over and above that, slight movements should be made with the controlling finger to produce the rises and falls in pitch.

At the suggestion of Dr. Damsté a design for a pitch controlled EL was considered in which clenching the fist would raise the pitch, gradual release would produce the declination and fingertip-pressure would provide pitch accentuation. However, we felt that this construction would not allow a stable application of the EL to the neck thereby increasing the chance of a sound leak through incomplete contact. Secondly, for reasons mentioned earlier in this section, the declination control should not be left to the EL speaker. The same line of reasoning applies to the continuous nature of the pitch control.

It was decided to have a prototype built (by C.G. van den Bergh, technician of the Institute of Phonetics, Utrecht) which incorporated the following characteristics. First of all a fixed declination function was specified: a gradual lowering of the overall pitch at a rate of -2 ST/s (this was changed later to -1.5 ST/s). In experiments on synthesizing pitch patterns at the IPO (Eindhoven) this value had been established empirically as suitable for reasonably short British English utterances (De Pijper, personal communication) (for Dutch this value eventually proved to be somewhat too high, hence the change).

Secondly, it was decided to incorporate a fast rise (comprising rises 1, 2 and 3, cf. section 1.2.3) and a fast fall (comprising falls A, B and C), thus facilitating the production of the hat pattern and its derivations. 't Hart & Cohen (1973) had specified the standardized fast rise at 4 ST in 100 ms and the standardized fast fall at -5 ST in 75 ms. For reasons of simplicity in the circuit design, however, as a fast fall a decrease in pitch of -4 ST in 100 ms was selected (Collier, 1972, stated that the slope of pitch falls is not very critical in speech synthesis). In this way it was possible to utilize a second level, 4 ST higher and parallel to the (lower) declination level. The desired pitch movements could then be realized by electronic control of the transition time (viz. 100 ms) from one level to another. These specifications resulted in a circuit of which an early version is described in Van Geel & Van der Werff (1980) and the most recent version in Cohen, Van den Broecke & Van Geel (to appear). The circuit causes the pitch to drop from 100 Hz levelling off to 65 Hz in 5 seconds if the lower declination is switched on permanently, and from 126 Hz to 82 Hz with the upper declination switched on permanently.

The pitch movements are made in the following way: the lower declination starts when the control button on the EL prototype is depressed. When it is slid forward pitch is raised towards the upper declination level and follows it. Pitch drops back to the lower declination level in 100 ms when the control is moved backwards again. Releasing the control either in the lower or upper declination level position causes the output to be switched off and the declination to be reset in silence at the beginning of the lower level. This was done deliberately as most Dutch utterances start at the lower declination level.

At the time when this study took place only two prototypes of the kind described above were available, which imposed another restriction of a practical nature on the number of EL speakers who partook in the learning program.

The specifications of the EL prototype used, and the reduced number of pitch movements which can be produced with it largely define the contents of the initial stage of the learning program, i.e. the selection of the chapters and sections in the intonation course used (Collier & 't Hart, 1981), cf. section 3.2.

The reduction of the number of pitch movements was made with two purposes in mind: to keep the pitch control as simple as possible and to keep the amount of theoretical knowledge necessary for adequate use of the device as limited as possible. The consequence of this reduction is that only hat patterns and derivations can be produced with the prototype EL. However, in about 70% of all Dutch utterances pitch patterns have been observed (cf. section 1.4) which consist only of a succession of fast rises and fast falls. The simple hat pattern in itself was found in about 33% of these utterances. One of the claims made by Cohen, Collier & 't Hart in their reports is that pitch patterns are interchangeable to some extent, on the condition that the number of pitch accents remains the same. This implies that pitch patterns with less frequent pitch movements can be replaced by hat patterns and their derivations. This claim is even made stronger by the observation in Collier & 't Hart (1981) that speech with only these pitch patterns will be perfectly natural to the casual Dutch listener.

As outlined in section 2.7 in the next chapter an intonation training program will be described together with its application. In this program the prototype EL was utilized exclusively; it was decided not use the Western Electric EL for reasons given above.

Chapter 3

DEVELOPMENT OF AN INTONATION LEARNING PROGRAM

3.1 Setup of a learning program for laryngeal speakers

3.1.1 Introduction

The working hypothesis for this study was: conscious knowledge of the perceptual and productional aspects of pitch inflection facilitates adequate use of the pitch inflection control on an EL. The crux of this hypothesis is the conscious knowledge. This points in the direction of a cognitive approach to teaching intonation production in EL speech.

It has been assumed in the various reports on the intonation of Dutch by Cohen, Collier, and 't Hart that perceptually relevant pitch movements, which are the basis of adequate pitch inflection, are the result of voluntary muscle movements. This was later substantiated by Collier (1975). It implies that pitch movements are subject to conscious control by the speaker, and that, once the speaker knows what to listen for and how these movements are described, they can be produced by him at will and when asked for.

The 'Cursus Nederlandse Intonatie' (Course in Dutch Intonation) by Collier & 't Hart (1981) is based on such an approach. It starts with training in listening to pitch movements which differ in direction, excursion, and place. In other words, the listener has to acquire conscious perceptual knowledge of pitch movements. Once this has been achieved the course continues with the relation between intonation and accentuation. Subsequently the production of pitch movements is dealt with. It is clear that this approach aims at providing conscious knowledge of pitch phenomena. This knowledge is then to be used as the basis for the production of correct pitch patterns (see also section 1.5).

A similar approach was used in the setup of a learning program for speakers who use an EL with intonation facilities which will be described in the following sections. A preliminary (1977) and unpublished version of the intonation course, though very similar to the published version of 1981, was used for this learning program.

By way of pilot study, and in view of the limited availability of EL speakers, the first setup of the learning program was meant for laryngeal speakers only. This was done to minimize the risk of confronting EL speakers with an improperly working learning program. The first participants were mainly phonetically trained persons.

In each step the participants were required to master the training task to the experimenter's satisfaction before continuing with the next step. This informal method was utilized as it did not appear justified to set criteria which were not based on experience. Further investigation will hopefully yield more explicit criteria.

3.1.2 Conscious knowledge of pitch phenomena

The participants were presented with the first four chapters of the intonation course mentioned above. Although the phoneticians in the group of participants could be assumed to have a basic knowledge of Dutch intonation theory, having them study these chapters ensured a comparable level of knowledge among all participants. The taped demonstrations accompanying the first chapter served to help the reader develop an ear for pitch movements in speech. The second chapter deals with typical Dutch pitch movements in running speech. The third describes simple intonation patterns built from these pitch movements and their place in the syllables. The fourth chapter introduces more complex patterns with the same pitch movements.

It was supposed that after having studied the chapters and having listened to the taperecorded demonstrations accompanying the course the participants were conscious of the perceptual and the productional aspects of their intonation.

3.1.3 Laryngeal imitation of pitch patterns

To consolidate the conscious knowledge of pitch production the participants were asked to imitate, with their own voice, seventeen model sentences which had been produced with the natural voice of the experimenter and had been recorded on tape. The FO of each sentence was measured and displayed on a screen by means of an F-J Electronics Pitch Meter and ditto Curve Display. To this aim a pitch equivalent audio signal had previously been derived from an electroglottographic (F-J Electronics) signal and recorded together with the microphone signal of the samples serving as models. This setup is shown in fig. 3.1.

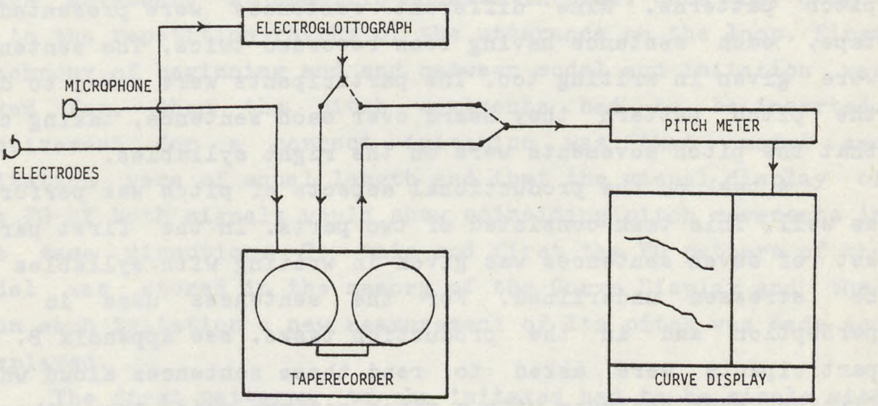


Fig. 3.1: setup for recording, measuring, and displaying FO.

The FO of the imitations was measured by means of the electroglottograph and displayed directly on the Curve Display. In this way the pitch contours of model and imitation could be displayed simultaneously. This is considered one of the main prerequisites for visual feedback (Abberton & Fourcin, 1975).

The experimenter's decision for a correct imitation depended on auditory and visual observation: it was checked visually whether the imitation contained essentially the same

F0 pattern as the model. Auditorily it was checked whether the pitch movements were on the same syllable. The model sentences themselves were taken from the intonation course and were, in fact, observed samples of everyday speech (see app. A).

After having gone through this stage in the program, the participants should be able to produce pitch patterns fitting to sentences given in writing (cf. section 3.1.4).

3.1.4 Tasks

To find out whether the participants were familiar enough with the pitch patterns introduced in the first stage of the program (3.1.2), they did a task in which they had to recognize pitch patterns. Nine different sentences were presented on tape, each sentence having been recorded twice. The sentences were given in writing too. The participants were asked to draw the pitch pattern they heard over each sentence, taking care that the pitch movements were on the right syllables.

A task on the productional aspects of pitch was performed as well. This task consisted of two parts. In the first part a set of seven sentences was given in writing with syllables to be stressed underlined. For the sentences used in the perception and in the production tasks, see appendix B. The participants were asked to read these sentences aloud while fitting each sentence with a pitch pattern which they thought was appropriate. In the second part the same set of sentences was presented together with complete F0 patterns in writing. These were taken from the intonation course. Again the sentences had to be read aloud. The probable occurrence of discrepancies between the patterns chosen by the participants were utilized to show them that more than one pattern can be fitted to the same sentence.

Both tasks also served to round off that part of the learning program in which the participants were made conscious of their own intonation, and to check the effectiveness of this program.

3.1.5 Manual imitation of pitch patterns with an EL

For this stage the goal was that the participants should acquire a certain degree of dexterity in handling the pitch controlled EL. The training task was to imitate the pitch of an model sentence almost synchronously with the presentation of that sentence; the EL was to emit sound into free air. The models had been spoken with the pitch controlled EL. They had beselected by a listening pnael from a set of recordings, the selection condition being that the timing of the movements was perceived as natural. Each model was presented on a tape loop to enable the participants to have an a priori, an instantaneous, and an a posteriori reference for the imitation. The required synchrony forced the participants to make the pitch movements at specific moments in time, once they caught on to the repetition rhythm of the utterance on the loop. First synchrony of beginning and end between model and imitation was asked for, then the pitch movements had to be inserted. Requirement for a correct imitation was that model and imitation were of equal length and that the visual display of the FO of both signals would show coinciding pitch movements in the same direction. To this end first the FO pattern of the model was stored in the memory of the Curve Display and then upon each imitation a new measurement of its pitch was made and displayed.

The first patterns to be imitated had to be simple with only one pitch movement; then more complex pitch patterns with two or three movements, in which consecutive movements followed each other more closely in order to increase dexterity on the part of the subject.

3.1.6 Instructions for use of an EL by normal speakers

The participants were acquainted with a monotonous EL (Servox). This was planned as follows: first the changes in the anatomy of the postlaryngectomy patient were explained, especially in relation to speech production. Then an EL was presented and the optimum spot on the neck had to be found with each participant for optimal transmission of the vibrations emitted from the device into the speech tract, i.e. with the least possible damping and minimal direct radiation.

Then a brief training followed to keep the glottis closed while speaking in order to simulate the anatomy of the laryngectomee (cf. Weiss, Yeni-Komshian, and Heinz, 1979) and to improve the artificial voice output level. With an open glottis too much energy gets lost in the tracheal and respiratory cavities. Keeping the glottis closed was effected by asking the participant to inhale, to close the glottis, and to exert slight subglottal pressure which could be kept up for a couple of seconds, long enough for the production of a sentence.

Limited training in the production of vowels and consonants followed. The difference between voiced and and unvoiced consonants (see also section 2.4) had to be realized by means of differences in buccal air pressure caused by extra articulation effort in the case of unvoiced consonants (cf. Isshiki & Tanabe, 1972). An exercise (adapted from Gardner, 1971: p.174) to achieve this was to say: 'pabapabapaba....', 'tadatadatada...', etc. first in pseudowhispered speech with closed glottis and then with the EL switched on.

The goal of this stage in the learning program was a usable level of intelligibility of the resulting EL speech.

3.1.7 Combination of skills

In the final stage of this learning program for normal speakers the EL speech production skill and the manual pitch production skill were combined when, by means of the EL prototype, the participants were asked to read aloud sentences presented in writing with given pitch patterns. The most simple sentences had to be tried first. The set of seventeen sentences mentioned in section 3.1.3 was used here too (see also app. A).

It was intended that each model had to be imitated correctly two or three times to the expeimenter's satisfaction, depending on the number of attempts preceding the first correct imitation; when a relatively high number of attempts was necessary, three correct imitations were asked.

The same tape loops were used as in the manual imitation stage (section 3.1.5) to alleviate this task of combining skills.

The whole program is given schematically in fig. 3.2.

CONSCIOUS KNOWLEDGE

TASK

LARYNGEAL IMITATION

2 TASKS

MANUAL IMITATION

EL INSTRUCTION

READING ALOUD +
PATTERNS

Fig. 3.2: diagram of the stages of the learning program for normal speakers.

3.2 First application of the learning program

3.2.1 Participants

As mentioned earlier the participants were mainly phonetically trained persons. Five were colleague-phoneticians. One of the two remaining participants was not phonetically trained, the other was a student of phonetics. These two were included to try and discover gaps in the program due to phonetic knowledge taken for granted, since the eventual laryngectomised participants could not be supposed to have any previous knowledge of the phonetic theory presented in the program. The phoneticians served a triple role: that of learners of intonation production, that of critics of the setup of the course and that of quick learners thanks to prior knowledge of intonation theory. With this construction, it was hoped, major flaws in the program were eliminated before application to alaryngeal speakers.

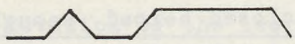
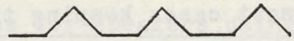
3.2.2. Procedure

The participants required about 45 minutes to study the chapters of the intonation course and listen to the accompanying demonstrations. A listening task in chapter one was performed in which twelve pairs of pitch movements on a vowel differing in direction, excursion, duration and place were presented on tape and had to be identified as to the characteristics of one of these four parameters. All participants scored better than chance (less than six wrong choices). The part in chapter three dealing with the place of a pitch movement in a syllable was judged too early in the program. Chapter four was found to give too much information on pitch movements which could not be made with the EL prototype.

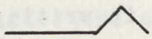
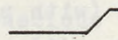
Reproducing the pitch patterns did not cause many problems: most participants required about 20 minutes to finish the seventeen sentences. In some cases where a participant stated that the pitch patterns prescribed there would not have been his choice, relatively more time was needed than for other sentences. Some participants introduced pauses, which resulted in other pitch patterns than intended. These patterns often included continuation rises before the pauses. Participants would have to be asked not to make any pauses where none were indicated.

In the perception task the participants were asked to use the first presentation of each sentence to draw the pattern heard and the second presentation for checking what had been drawn and, if necessary, correcting it. Most participants needed about 12 minutes for the 9 sentences.

An error often made was that the following pattern (A) was heard as pattern (B).

(A)  (B) 

The participant who was not phonetically trained displayed the following tendencies in his responses: he did not perceive all pitch accents, and instead of pattern (C) he often heard (D).

(C)  (D) 

These errors may have been caused by insufficient skill in recognizing pitch patterns.

In the pattern production task the participants showed the same tendency as in the pattern imitation task: they introduced pauses both when the participants were free to choose their own pitch pattern and when the pitch patterns were predetermined. In the latter case, obviously, making pauses was considered a mistake, since no pauses had been indicated. Instructions when to make or when not to make pauses would have been useful here: also sentences including pauses ought to have been present in the learning program. Leaving the choice of the pitch pattern to the participants indeed facilitated the production of the pitch patterns.

The participants needed two sessions for manual pitch pattern imitation of 17 sentences. Despite the short interval between the end of one and the beginning of a following presentation of the model sentence (which had been recorded on a tape loop) the participants found it hard to pick up the rhythm at the moment of switching on at the beginning of the sentence. The participants had to give two or three correct imitations for each sentence, without having to synchronise the beginnings of the sentences perfectly, for the above mentioned reason.

The instructions for the use of an EL were easy for all participants. The voiced/voiceless opposition in consonants could indeed be made by means of different buccal air pressures. Keeping the glottis closed was also easily learnt, although imitation of consecutively pronounced sentences caused many participants to run out of breath. To find the correct spot for placing the EL against the neck was not always easy, but in most cases keeping the glottis closed helped enough to render the speech output to be clearly intelligible.

The combination of the skills of EL speech production and manual production was made by two persons. They did this successfully according to the experimenter. After they had imitated the seventeen sentences from tape loops they were able to read them aloud (with pitch patterns given in writing). The other participants did not go through this stage of the program, since it was felt that enough information had been gained for making the necessary revisions to the learning program in order to be able to apply it to laryngectomized persons.

3.2.3. Conclusions

The comments from the participants and the way they performed indicated that the following changes in the learning program were necessary:

1) An excerpt limited to information relevant for the EL prototype of the first chapters of the 'Cursus Nederlandse Intonatie' should be used and presented orally to the participants. In this way of presentation superfluous information would be left out, and it was hoped that it would be easier to check whether all the information presented in these chapters was understood by the participants.

2) It should be stressed that no pauses should be made unless indicated. By implication, sentences with pauses would have to be added to the training material.

3) A perception training should be inserted into the program, roughly similar to the subsequent perception task; both should precede the production training and task. In the perception training the pattern drawn by the participant should be compared with that of the model given on the curve display.

3.3 Revision of the learning program for speakers using an electrolarynx

Apart from implementing the first two requirements as given in 3.2.3. the following stages of the learning program were changed to comply with the third requirement given above.

After the stage in which conscious knowledge of pitch phenomena was acquired perception training took place. Of each of the 17 sentences pitch was measured and displayed after a participant had drawn the pattern he heard. As before, conditions for a correct pattern were: same pattern as the model, and pitch movements indicated on the same syllable as in the model. After having completed this stage the perception task as described in 3.1.4 was planned.

The stage following the perception task dealt with production of pitch patterns as the laryngectomees used to produce before their operation. This means that the laryngectomized participant had to become conscious of the mental activity of producing pitch variations. Since the source for the production of pitch was the larynx, a different solution had to be found. It was decided to familiarize the participants with the Memacon DSP8 pneumatic artificial larynx (for a description of its working principle see Lebrun, 1973). With that device the user can control the pitch of the voice produced to a limited extent by varying exhalation pressure. In this way an almost natural way of pitch control is effected. It was hoped that a brief training in speech production with this device would enable the participants to speak and intonate with it. The production task, then, could be the same as the one used in the learning program for laryngeal speakers.

3.4 Additional steps towards spontaneously intonated EL speech

3.4.1. Introduction

There is still a large gap to be bridged between reading aloud sentences and spontaneously intonated EL speech with given pitch patterns, which was the skill acquired at the end of the program described so far. It was thought that the following steps would gradually lead the participants towards that goal. With each step less information was given about the precise form of the pitch pattern and more was required from the participant.

3.4.2 Reading aloud without given pitch patterns.

The first step towards spontaneously intonated EL speech was to have the participants read sentences with only the place of the pitch accents indicated by means of underlined syllables. At this moment in the learning program the participants are supposed to have enough insight in the intonational possibilities to be able to fit a pitch pattern to the accents indicated. The participants drew the pitch pattern and the experimenter commented upon it, approving or correcting it, as the case might be, before the participant was asked to read the sentence with the chosen pitch pattern aloud.

The training material was a set of simple sentences which increased in length, so that the first sentences mostly contained one or two pitch accents and later sentences contained up to four or five pitch accents.

Testing whether a pitch pattern had been chosen and produced correctly was hardly possible. The only really reliable way would have been to run listening tests utilizing judging panels. The training situation did not allow such indirect methods, as direct feedback was required. Therefore the experimenter passed judgments on the correctness of the intonation of the sentences produced to the best of his ability. This way of testing was also adopted in following stages.

Then the sentences were presented without any indication about pitch or accentuation. The participants were asked first to indicate the pitch accents which were then discussed as to number and place. After approval or correction by the experimenter a pitch pattern had to be fitted, drawn and duced. Two or three correct productions as judged by the experimenter were required per sentence.

3.4.3 Question/answer game

Once reading aloud of sentences without help about the intonation was mastered, the next step consisted of participation in a question/answer game. The learner was required to answer a simple everyday question like 'What day is today?', 'How is the weather?', 'Have you been out already?' etc. The answer had to be simple but complete, the participant

CONSCIOUS KNOWLEDGE

TASK

PERCEPTION TRAINING

TASK

PRODUCTION TRAINING

TASK

EL PROTOTYPE INSTRUCTION
+ MANUAL IMITATION

READING ALOUD
+ PATTERNS

SAME
+ ACCENTS

SAME
- ACCENTS

QUESTION/ANSWER
GAMES

SPONTANEOUS EL
SPEECH + INTONATION

Fig. 3.3: diagram of the stages of the learning program for EL speakers.

concentrating not on the content of his answer, but on its form, especially its intonation. The participant was instructed to think first how his answer would sound, how the intonation would be, and then to say it aloud. He was asked to continually listen to himself to check how he intonated. Gradually the questions became a little bit more personal, thus eliciting spontaneous remarks between answers. In this way the participant would slowly progress from the skill of reading aloud to spontaneously intonated EL speech.

A diagram with brief descriptions of the stages of the learning program resulting from the changes and additions described above is given in fig. 3.3. Results of the procedure adopted and outlined here will be presented in the form of the following case histories.

3.5 Application of the learning program with EL speakers

3.5.1 Case study of speaker A

When the first participant started to follow the program he was 57 years old. Before his laryngectomy he had been executive manager in the training section of the Dutch national railways. Audiometric screening showed that this person did not have any hearing deficiency. He was considered a superior EL speaker.

The first stage of the learning program in which the participant was acquainted with pitch and intonation theory took 1 1/2 sessions of about 75 minutes. In the task on discriminating between two different pitch movements seven out of eight questions were scored correctly.

The perception training stage also took 1 1/2 sessions of about 50 minutes. In the perception task four of the sixteen pitch accents in the nine sentences were missed, which was about equal to the average performance of the laryngeal participants with this task. Both tasks showed that the aims of the corresponding stages in the learning program had been fulfilled.

Two sessions were spent on learning to speak with the DSP 8 pneumatic artificial larynx. Due to the oblique form of the stoma of this speaker it was hard to maintain an airtight seal between the skin and the device. The output of the few sentences that came out well could not be analyzed as to their F0 with reliable results, thus precluding visual feedback of the pitch contour. It was decided to skip this stage as spending more effort and time on mastering speech with the device would distract from the real purpose of the learning program. Then the EL prototype was introduced and the operation of the pitch control button was explained. As the pitch control button had to be pushed and slid, the thumb appeared the best choice in this case; the participant was advised accordingly and in practice the choice proved to be justified.

The seventeen sentences originally planned for the production stage (see app. A) were used in this combination stage. It was hoped that dexterity in handling the EL prototype would come with training in the imitation of the example sentences. With every next sentence a correct imitation, according to the experimenter's judgement, was attained in fewer trials. In some apparently difficult cases visual feedback of the F0 contours of both example and imitation was employed. Sentences with consecutive rises and falls were ordered in such a way that these pitch movements followed each other more closely with each new sentence, until a rise and a fall combined into one peak. This required the subject to move the pitch control button very quickly.

The participant was asked to take the EL prototype home to practise. In the seventh session already, while still training on reading aloud the seventeen sentences with given pitch patterns, the first spontaneous remarks were made with acceptable pitch patterns. Then the same set of sentences was given with only the places of pitch accents indicated. In a next session a set of rules for selecting a suitable pitch pattern was presented to aid in drawing a pitch pattern. The reason for offering these rules and a description of them are given in section 3.5.2. From now on a set of sentences of increasing length was used to be read aloud. After a correct pitch pattern had been fitted the sentence was read aloud. The procedure was repeated with sentences without any given pitch accents; it was left to the participant to place the accents. Gradually less time was spent on actually drawing pitch patterns. Then the patterns had to be constructed mentally before the sentences were read aloud.

In the tenth and subsequent sessions question/answer games were done next to practising in reading aloud. Reading a piece of continuous text (the Dutch version of the fable of the north wind and the sun as given in IPA, 1949) proved to be difficult, mainly because the sentences in that text were fairly long.

The eighteenth session was the first session completely filled with spontaneously intonated EL speech. Subsequent sessions were held similarly. In the twentieth session a new version of the EL prototype, consisting of one single unit, was given to take home.

In total 40 sessions were held with this participant of which the first 18 dealt mainly with learning a new speech mode and the remaining sessions served as training. The learning program covered those first 18 sessions.

3.5.2 Case study of speaker B

This subject was 86 years old when he started to follow the learning program and was moderately hard of hearing, considering his age. Before retirement he had been a railway employee and an active member of the trade union. He did not wear a hearing aid. Auditory tests indicated that he suffered a hearing loss in the upper frequency region. Mean centre octave hearing loss for pure tones was 30 and 35 dB for right and left ear respectively. Phoneme discrimination loss was also observed.

Presumably, his affected hearing hampered him in listening to pitch movements, especially when doing the task at the end of the first chapter of the intonation course (only 25% correct). Furthermore, the part on intonation theory proved too abstract for him. For these two reasons it was decided to present a set of simple rules and a list of alternative patterns for each number of pitch accents to be made in a sentence (see appendix C). This set of rules was also presented to the first participant, enabling him to choose a pitch pattern fitting to the number of pitch accents of a given sentence. In order to prepare the participant for this task he was trained to try and listen for pitch accents as early as in the perception training stage. By means of this set of rules we made an analysis and drew a pitch pattern of each sentence of the set of seventeen (cf. section 3.3). Especially the rule stating that most patterns started and ended low proved to be a great help.

Producing sentences with given pitch contours or pitch accents was very difficult for him. He complained that the sentences were too distant from his own language. He then was asked to bring sentences of his own (session 9). This appeared to be an improvement. These sentences and two more sets provided by the participant were used first with pitch patterns fitted by the experimenter, then only with given pitch accents. However, the improvement was relative and shortlived: over a period of four consecutive sessions this speaker did not show as much progress as had been expected. For a change a question/answer game was introduced (in session 13). Apparently, the near-spontaneous speech mode was what was needed for a breakthrough. From then on he progressed more quickly. The pitch movements in his own sentences, when read aloud, were more precisely placed. With every following session more time was spent on spontaneous speech.

As he sometimes forgot to intonate when talking about subjects he clearly had strong feelings about, he was reminded to constantly listen to himself to check whether and how he intonated. (This phenomenon reflects the complexity of the speech task: the EL speaker has to pay attention not only to <u>what</u> he says, but also to <u>how</u> he says it.)

In session 30 we had reached our goal: relapses had decreased to a minimum. In following training sessions he occasionally produced a sentence without intonation, but he was always notified by the experimenter of such a relapse. In spite of this, he never really managed to overcome this lack of concentration.

3.5.3 Case study of speaker C

When the third participant started to follow the learning program he was 76 years old. He had upper frequency hearing loss above 1000 Hz. Speech recognition 25% at 25 dB and 100% at 60 dB. Recruitment was found for the left ear: words presented at 60 dB were understood for 80%, at higher levels of presentation this percentage decreased.

Beginning with the first session the last five minutes of each session were spent on speaking with the EL prototype. In the first session the pitch control button was explained. He was asked to speak with the EL prototype while trying to use the pitch control without putting too much emphasis on producing completely correct pitch patterns.

The task on differentiating between two pitch movements was scored 75% correct. But when all intonation theory had been explained the participant still did not grasp the connection between controlling the EL prototype and the intonation theory. The use of visual feedback also posed a problem. He did not quite understand how the visual representation of a pitch contour could be linked to the auditory reality. He could not work with this different mode of presentation.

In the perception training stage he showed the same behaviour. However, he was able to imitate the pitch patterns he heard and knew that what he imitated was correct. Extra training on drawing pitch patterns temporarily helped but in the perception task he again drew inappropriate patterns. Yet, when asked to imitate what he heard, he made the same patterns as were presented. The production training was started with explaining the set of pitch production rules. Especially the diagrams with the alternative patterns for each number of pitch accents per sentence seemed to work well with this speaker. Soon he was able to read the training sentences without having to work out a pitch pattern beforehand.

After four sessions of question/answer games the speaker received the EL prototype to take home with him. However, being a widower, he did not have much chance of talking with the device at home, and he was not a very talkative person anyway. Lack of practice impeded his progress.

It was hoped that he would soon be able to use the EL more freely, but apparently this was not the case since he even received some negative comment on his new kind of speech. With a third person listening in on a session he was rather agitated, more than had appeared in sessions with only the experimenter present. This behaviour might have been present when he had been talking to others, thus deteriorating his performance. He stated that he found speaking with the device rather tiresome, which is not surprising considering the complexity of this mode of speech. On the other hand he stated that he liked speaking with the device. In 21 sessions he nearly attained correctly intonated spontaneous EL speech, but had to discontinue the sessions due to illness.

3.5.4 Conclusions

The application of the learning program with three EL speakers yielded three important changes which were combined into a third version of the learning program, proposed at the end of this section. These changes will be described in short.

The first change consisted of leaving out pitch production training for which speech via a pneumatic artificial larynx had to be mastered. Producing speech with such a device appeared to be not as easy to master as EL speech, and the output was not very suitable to be measured as to its pitch. Secondly, the stage in which learning to control the pitch on the EL prototype and learning to read aloud sentences were combined, could be made easier when the introduction of the EL and the familiarization with it were moved to the first session in the program. The third change was the addition of a set of rules for fitting pitch patterns to sentences on the basis of the number of pitch accents to be made. In the stage in the program where intonation theory is dealt with, more attention should be paid to the notion pitch accent and the way a pitch pattern is linked with it (for the pitch pattern production rules see app. C). The list of alternative patterns for each number of pitch accents should appear in the stage where reading aloud was practised.

The third version of the learning program, based on the previous version and incorporating the abovementioned changes is given in fig. 3.4.

CONSCIOUS KNOWLEDGE
+ ACCENTS + RULES
+ EL PROT. INSTRUCTION

TASK

PERCEPTION TRAINING

TASK

READING ALOUD
+ PATTERNS

SAME + ACCENTS
+ ALTERNATIVES

SAME - ACCENTS

QUESTION/ANSWER
GAMES

SPONTANEOUS EL
SPEECH +INTONATION

Fig. 3.4: revised learning program for EL speakers.

In the next chapter an evaluation of the resulting intonated EL speech will be presented.

Chapter 4

ASSESSMENT OF THE QUALITY OF EL SPEECH PRODUCED WITH INTONATION

4.1 Naturalness of pitch in EL speech

4.1.1 Introduction

Once it had been established that controlling the pitch of an EL by hand is an activity that can be learnt and can be executed in spontaneous speech, it became important to find out where the advantages of intonated EL speech over monotonous EL speech lie. It was thought that the naturalness of EL speech would be enhanced especially for listeners unacquainted with this kind of speech. People often complain that monotonous EL speech is unpleasant to listen to. It is difficult for these listeners to concentrate on listening to the EL speaker. When confronted with intonated EL speech many stated that in this case the complaint did not apply. To what extent the naturalness was enhanced by intonation and which factors contributed most to it remained to be assessed. It was expected that the factor pitch movements was more important than the factor declination, as pitch movements can have a functional load with respect to the perceptual prominence of words, whereas declination can be considered the consequence of a physiological process (Collier, 1975) and, therefore, having no function as regards lending prominence to a word.

A test was set up to answer two questions. The first question was whether and to what extent correctly intonated EL speech would be judged as more natural than monotonous EL speech. The answer to this question would be an indication about the perceptual distance between monotonous and intonated EL speech. The second question was: which of the two factors 'pitch movements' and 'declination' contributed more to the naturalness of intonation patterns in EL speech? The outcome would show us to what extent the presence of declination influenced the perceived naturalness of the pitch patterns. As a consequence it might be decided to maintain or leave out the declination function in the EL prototype.

4.1.2 Method

To answer the first two questions of the previous section it was decided to test a set of sentences which appeared in various realizations. By way of reference a monotonous version of each sentence was included. To ensure that the quality of the source signal did not interfere with what was to be the only independent variable, viz. different realizations of pitch patterns, the monotonous versions were not recorded with a monotonous Servox EL, but with the same device wiith which the other versions were recorded.

Another version to be tested was, of course, one with a correct pitch pattern, as was defined in section 3.1.3. Between the extremes of no pitch pattern and a correct pitch pattern there were two 'stripped' versions. The first was a version with pitch movements, but without declination. This version was representative of the factor 'pitch movements'. The second was a version without pitch movements, but with declination. Similarly, this version could be considered representative of the factor 'declination'. Comparison of the judgements for both versions would yield the relative importance of the declination factor vs. the pitch movements factor.

The first subject to follow the learning program had become so proficient that he could produce sentences with many different correct or incorrect intonation patterns when they were given on paper. He was asked to read a set of seven sentences in four ways: monotonous (M), with pitch movements only (P), with declination only (D), and with both declination and pitch movements (complete and correct, C). The way in which the different versions were produced will be explained later in this section.

The judges for this test were naive listeners; they were unacquainted with EL speech. The judges were third year university students of Dutch who cooperated without a financial incentive. They followed a course in phonetics and could be expected to have a critical ear for speech characteristics.

A test was run in which the listening panel was asked to judge the naturalness of the intonation in the sentences presented on tape, and to indicate their judgements on a five point scale.

The outcome of this test was completely counterintuitive: the M versions were judged to be more natural than their corresponding C versions. After a second look at the test sentences, which had been fitted with a pitch pattern by the experimenter, it appeared that these sentences contained pitch patterns which, although not necessarily unnatural for these sentences, required extra context. Apparently sentences without extra context can only be accepted as natural when the most 'neutrally' placed pitch accents and the accompanying pitch patterns are chosen.

For these reasons the test was regarded as a pilot test for checking the stimulus material and a more extensive test was set up, by which it was hoped all the questions of section 4.1.1. could be answered. In addition, a version was included by which a third question might be answered. This question was whether incorrectly intonated EL speech would be judged as more natural than monotonous EL speech. The answer to this question would tell us something about the importance of the learning program with respect to training in the production of pitch patterns. An incorrect version was characterized by pitch movements on syllables, or even words, which should not be accented.

It was necessary to utilize simple sentences not requiring any extra context in this test in which the sentences were to be judged separately. It is hard to make the notion 'not requiring extra context' explicit, but the following examples will make this clear: 'they had been fitted with a pattern'. is a sentence that needs more context than: 'Each sentence was fitted with an intonation pattern'.

With considerations like these a selection was made of spontaneous sentences which had previously been recorded on tape during the training sessions in the application of the learning program. Of each of the seven sentences selected five other versions were recorded for the main test: so apart from the version spontaneous (S) there were complete and correct (C), monotonous (M), pitch movements only (P), declination only (D), and incorrectly intonated (I). The specifications of these versions are given in table I. Version C was similar to S, but for the fact that it was read aloud, as were all other versions (see column 'spontanous'). It was expected that versions S and C would not be judged as significantly different.

Table I: specification of the six versions of the sentences presented.

version	declination	movements	spontaneous	correct pattern
monotonous	−	−	−	(no pattern)
pitch movements	−	+	−	+
declination	+	−	−	(no pattern)
correct	+	+	−	+
spontaneous	+	+	+	+
incorrect	+	+	−	−

On the basis of observations in intermediate stages in the application of the learning program it was assumed that incorrect intonation is characterized mainly by pitch movements at the wrong places, which was considered the result of providing the EL user with instructions without training in their application.

The prototype EL used in the recordings of these different versions was equipped with a switch that bypassed the declination function of the device. The column 'declination' indicates whether the declination function was switched on or off. In the latter case pitch was set at 100 Hz, being the starting point in the lower declination line (see also section 2.4). For a list of the sentences with the various pitch patterns see appendix D.

Before presentation of the sentences the naive judges listened to a sample of monotonous speech of the same speaker to get accustomed to his manner of speech and to EL speech in a general sense. This time the judges were 12 third year students of French, all were native speakers of Dutch. They attended a course of phonetics, so they could be considered critical judges of speech characteristics. They were asked to indicate their judgements on a 5-point scale on which 1 stood for very unnatural, 5 for very natural and a 3 indicated a neutral judgement: neither natural nor unnatural. All sentences were given on paper in the hope to prevent possible difficulties in understanding what was said from clouding the listeners' judgements.

4.1.3 Results

The raw scores were values at an ordinal level of measurement, as it could not be assumed that the intervals between the points of the scale were equal for a single judge. By means of the method of successive intervals the scores were converted into scale values lying on an assumed psychological continuum (cf. Edwards, 1957). This continuum reflects the judges' appreciation of what is natural for pitch in EL speech. The resulting scale values were at ratio level. The mean scale values, calculated over the seven sentences, for each of the six versions are given in table II:

Table II: mean scale values of the six versions.

M: 0.48	P: 2.19
I: 1.10	C: 2.92
D: 1.20	S: 2.97

boundary value between natural and unnatural : 2.18

In table II it can be seen that the M and the S versions were almost at the extremes of the scale. Furthermore there was a large distance between the values of the left column and the right column. To find out whether these differences were significant the data were subjected to the following tests.

This naturalness test can be described as a single factor experiment with repeated measurements on the same elements. A oneway analysis of variance for this case is described in Winer (1971). The analysis resulted in $F(5,30)= 40.19$, $p<0.00001$. This indicates that there is a highly significant difference between the six versions. To find out which of the six versions are responsible for this difference a Scheffé post-hoc analysis was carried out on these means. The main difference was between the subgroups as divided by the naturalness boundary. Versions M, I and D did not differ significantly from each other, and for the same reason P, C and S could be grouped together.

4.1.4 Discussion

The mean scale values of C and S are similar, as could be expected. Version D, i.e. the 'declination factor' in versions C, S and I, scored below the boundary of naturalness. This means that this factor did not contribute as much as the factor 'pitch movements' did to the naturalness of the intonated sentences. The fact that only versions C and S were judged as natural indicates that the declination function should be maintained in the design of the EL prototype.

Furthermore the fact that version P did not differ significantly from C and S indicates that the factor 'pitch movements' contributed most to the naturalness of the intonated sentences. On the other hand version P scored almost the same mean scale value (2.187) as was the converted value for the boundary between natural and unnatural (2.184). This implies that, strictly speaking, only versions C and S can be considered as natural.

The judges scored version I only slightly, but not significantly higher than version M. The perceptual distance between correctly intonated EL speech and incorrectly intonated

EL speech was almost as large as the distance between correctly intonated EL speech and monotonous EL speech. By implication incorrect intonation can be considered just as unnatural as no intonation.

This stresses the importance of adequate training in producing pitch patterns. Without such training the risk of producing incorrect pitch patterns would be too great, thus nullifying all the efforts invested in following the whole learning program. Supplying the pitch controlled EL to a laryngectomee without having him follow the learning program would produce similarly negative results.

4.2 Naturalness of pitch and listening population

4.2.1 Listening population

Once it had been established that correctly (and spontaneously) intonated EL speech was distinctly more natural than monotonous EL speech or any other kind of intonated EL speech, it became interesting to know whether the difference in naturalness between intonated and monotonous EL speech was equally large for less naive listeners. It was decided to look at three other kinds of listeners: first year students of speech therapy, relatives of EL speakers (referred to as EL listeners), and EL speakers. The motivation for the choice of these particular listeners was the following: the speech therapy students are considered to have a trained ear for regularities and irregularities in speech because of their schooling. The EL listeners were chosen because of the fact that they hear EL speech everyday. Therefore they may be assumed susceptible to changes in the quality of EL speech. The EL speakers were included because they are the most important party in the communication process under study. To avoid possible bias in favour of intonated speech the three subjects who had followed the learning program were excluded from this group.

All three groups of listeners participated without financial stimulation. It can be added that the motivation of the EL speakers and EL listeners was extra high due to an obvious interest in EL speech quality.

Five EL speakers and five EL listeners partook in the test. Four of the listeners were wives of the EL speakers and one was a son of the only female EL speaker in the group. The number of EL speakers and listeners was limited by practical matters such as availability and auditory and visual impairments (remember the age of the average EL speaker!).The number of speech therapists was 29.

4.2.2 Method

Sentences were used which were spoken by the same subject as in the previous test. The sentences were also based on spontaneously uttered sentences which had been recorded previously. This time, however, ten sentences were used in three versions, viz. M, C, and P as an intermediate version. The choice of this particular version was motivated by the observation that the mean scale value for this version in the previous test was about halfway the M and the C versions (cf. section 4.1.3.). The sentences together with their pitch patterns appear in appendix E. Furthermore all stimuli were presented twice to facilitate testing for reliability of the judges.

The order of presentation was such that the sentences appeared in a Latin square design to balance out any ordering effects, as far as sentence number and version were concerned. The resulting series of 3 times 10 sentences was then repeated in reversed order of presentation avoiding immediate repetition of the 30th item by postponing its repetition to the very end of the complete series. The fact that the 30th stimulus did not appear two times consecutively was enough for the judges to fail to notice that the order of the second half of the test was almost a mirror image of the order of the first half. Again the sentences were given on paper; behind each sentence the judges could indicate their evaluation of its intonational quality by placing a mark in one of five squares, along the five point scale.

4.2.3 Results

The raw scores were converted into scale values by means of the method of successive intervals. The mean scale values (calculated over 2 x 10 stimuli per version) of the three versions for the three listening panels are given in table III together with the correlation coefficients between the scale values for the first and the second presentations of the stimuli.

Table III: mean scale values of the three versions for three listening panels, and statistics

	EL speakers	EL listeners	first year sp. therapists
M monotonous	1.35	1.42	0.69
P pitch movements only	1.62	1.36	2.36
C correct pitch pattern	2.03	1.79	3.38
naturalness boundary	1.35	1.05	2.39
Pearson corr. r=	0.69	-0.43	0.95
p<	0.005	0.01	0.005

In this table the boundaries of naturalness are the converted values for the '3's of the five point scales as used by the listening panels.

As can be seen in the table the panel of EL listeners was not reliable in its judgements of naturalness. A possible reason will be given in section 4.2.5. The other two panels showed a moderate to high correlation between the judgements of the first and the second presentation of the stimuli. Therefore it appeared justified to use the average value of both presentations with these two panels in the calculation of the means.

A oneway analysis for repeated measurements was carried out on the scale values of the two selected panels. This resulted in the following F ratios. For the speech therapists: $F(2,18)= 280.80$, $p< 0.0001$, and for the EL speakers: $F(2.18)= 12.49$, $p< 0.01$. A Scheffé post hoc analysis showed that for the EL speakers the versions could be divided into two subgroups: the M and the P versions together on the one hand and the C version on the other. In the case of the speech therapists all three versions differed significantly from each other.

4.2.4 Discussion

In the case of the EL speakers and of the speech therapists the version with correct pitch patterns was judged as natural. With the EL speakers there even was a division into two subgroups, viz. C vs. M and P. With the speech therapists all three versions differed significantly but only the C version was judged as natural, i.e. was scored higher than the boundary of naturalness (see table III).

The EL listeners panel produced counterintuitive judgements on the three versions: the M versions had a higher mean scale value than the P version. The correlation coefficient, however, between the scale values belonging to the first and the second presentation of the stimuli was negative and significant. This can only be explained by the presence of an artefact which played a role in the way the EL listeners gave their judgements. It may possibly be assumed that they somehow used the manner of speaking of their husbands, or mother, as a reference for naturalness instead of their intuitions about naturalness of pitch in general. The high level of significance of the negative correlation may have been caused by a combination of two factors, viz. the limited number of judges and the artefact which possibly produced the counterintuitive mean scale values. Considering the high correlation coefficient of the judgements of the EL speakers which were equally limited in number it seems justified to say that the abovementioned artefact was the main cause for the significant negative correlation coefficient found with the EL listeners.

The EL speakers judged all versions as natural. Even the M version was scored marginally higher than the naturalness boundary (resp. 1.350 and 1.345). This might have been caused by another artefact. The EL speakers commented that they found the sound of the device used in the production of the stimuli more agreeable than the sound of their own EL (both devices used the same vibrator, but these were driven by signals having different waveshapes, cf. section 2.4). This artefact, apparently, did not interfere with the relative ordering of the versions.

4.3 Conclusions

In a general sense it can be said that only EL speech with correct pitch patterns was judged as having natural pitch. Furthermore it appeared that pitch movements made the greatest contribution to this naturalness, although the factor declination was important for naturalness in the sense that sentences with only pitch movements and no declination were scored significantly lower than sentences with complete (and correct) intonation. Incorrect intonation was considered just as unnatural as no intonation, which stresses the importance of adequate training in the production of inflected EL speech.

4.4 Intelligibility of intonated EL speech

4.4.1 Introduction

Almost the same questions that were asked with respect to the naturalness of the pitch of EL speech could be asked with respect to the intelligibility of intonated EL speech. In this case the questions were, in short: can intonated EL speech be understood more easily than monotonous EL speech, does the factor pitch movements have more influence on the intelligibility of intonated EL speech than the factor declination, and is incorrectly intonated EL speech easier to understand than monotonous EL speech?

4.4.2 Method

All stimuli were spoken by speaker A. A pilot test had shown that the intelligibility of the speech of this EL speaker was very high. For this reason it was decided to add pink noise to the sentences in order to lower the intelligibility score so as to avoid ceiling effects. The noise level was modulated by the amplitude envelope of the speech signal (time constant 25ms). In another pilot test a speech/noise ratio of +5dB yielded an average intelligibility level of 66%. Natural speech presented at +5dB S/N was understood for more than 70% and at -5dB S/N for less than 50% (Young & Harbert, 1970). This implies that the EL speech of speaker A did not score appreciably lower than natural speech in these conditions. With this in mind an S/N ratio was chosen of 0dB in order to attain an intelligibility level of about 50% (cf. Denes & Pinson, 1973). In this way, it was hoped, bottom and ceiling effects would be avoided.

A set of ten sentences was recorded in five versions: monotonous (M) pitch movements only (P), declination only (D), correct pitch pattern (C), and incorrect pitch pattern (I). For the sentences and the pitch patterns used see appendix F. As described in section 4.1.2 sentences were selected which did not require any context.

It is obvious that a sentence could only be presented once to the same listener, as the sentence would be too easy to understand upon second presentation. Yet the intention was to use the abovementioned five versions for each sentence. Therefore a design was made in which five groups of listeners (all were pre-candidate Dutch students of English) containing

15 persons each were presented with five different stimulus tapes. Each group listened to all 10 sentences, but the sentences were presented in different versions to each group of listeners. This also implies that group heard different sentences for each version. Furthermore, each version appeared twice on each tape. The order of the sentences on each tape was such that two equal versions never appeared consecutively.

The listeners were asked to write down anything that they heard, even parts of words, as the number of correctly heard syllables was counted.

4.4.3 Results

For purposes of comparison the scores for each sentence in each version were converted into percentages of the total

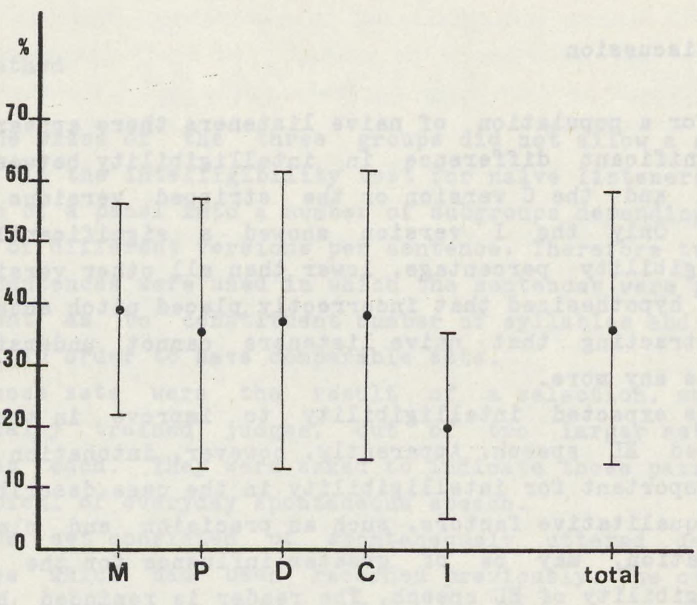


Fig. 4.1: mean percentages of correctly perceived syllables and SD's for the five versions.

number of syllables per sentence, as the sentences had different numbers of syllables.

The five groups all came from the same population (cf. 4.4.2). Therefore it appeared justified to combine the five groups into one listening panel. The mean values for each version and their standard deviations are depicted in fig. 4.1.

The low overall intelligibility percentages were not expected (cf. 4.4.2.). For reasons of comparison, however, the same S/N ratio was used in the experiment described in 4.5.2.

Figure 1 shows that only version I was scored lower than the others. This apparent difference was subjected to statistical testing to find out about its significance. Oneway analysis of variance for repeated measurements yielded $F(3.36)=3.22$, $p<0.05$. A post hoc Scheffé test on the means of the versions showed that the significance of the outcome of the oneway analysis was due to the difference between version I and the other four versions.

4.4.4 Discussion

For a population of naive listeners there appeared to be no significant difference in intelligibility between the M version and the C version or the stripped versions of the latter. Only the I version showed a significantly lower intelligibility percentage, lower than all other versions. It may be hypothesized that incorrectly placed pitch accents are so distracting that naive listeners cannot understand the sentence any more.

We expected intelligibility to improve in the case of intonated EL speech. Apparently, however, intonation is not very important for intelligibility in the case described here. Other qualitative factors, such as precision and clarity of articulation, may be of greater influence for the absolute intelligibility of EL speech. The reader is reminded here that the stimulus material was produced by one highly rated EL speaker, which may account for the almost equal intelligibility percentages in four of the versions, i.e. all but version I.

4.5 Intelligibility of intonated EL speech for different listening populations

4.5.1 Listening populations

To find out whether less naive listeners also failed to use the intonation cue as a help in understanding EL speech it was decided to try and find out what the possible difference was between monotonous and correctly intonated EL speech in the case of more or less initiated listening populations. As mentioned in section 4.2.1, and for the same reasons, three different groups of listeners were chosen: the sizes of the three groups were the same as in the test for naturalness, viz. 29, 5 and 5, respectively. (The EL speakers and listeners were visited at home for the naturalness test and the intelligibility test.)

4.5.2 Method

The sizes of the three groups did not allow a setup as applied in the intelligibility test for naive listeners, viz. division of a panel into a number of subgroups depending on the number of different versions per sentence. Therefore two sets of 10 sentences were used in which the sentences were pairwise equivalent as to constituent number of syllables and stress pattern, in order to have comparable sets.

These sets were the result of a selection, made by 6 phonetically trained judges, out of two larger sets of 19 sentences each. They were asked to indicate those pairs which were typical of everyday spontaneous speech.

One set consisted of spontaneously uttered intonated sentences which had been recorded previously. The other set contained monotonous sentences. For both sets see appendix G.

The sentences were presented in quasi random order: consecutively presented sentences always belonged to different versions and were never a related pair. All sentences were mixed with pink noise at OdB S/N ratio. The motivation for the choice of this ratio was described in section 4.4.2.

4.5.3 Results

For each sentence the percentage of intelligibility was derived from the number of correctly perceived syllables. Then for both versions the mean percentages were calculated. The percentages of C were higher than of M, but table IV shows that the standard deviations on those percentages were large. To find the real distance between M and C a Mann Whitney U-test was applied.

Table IV: mean percentages of intelligibility and outcome of Mann-Whitney tests for three listening groups.

	M		C	
	$\overline{x}$	SD	$\overline{x}$	SD
EL speakers	9.5	11.4	22.6	17.2
EL listeners	15.9	19.2	30.4	28.1
speech therapists	3.9	5.6	26.3	19.0

4.5.4 Discussion

It appeared that only for the speech therapists the difference was significant ($p<0.001$), for the EL speakers and EL listeners the differences were not significant. Apparently the high standard deviations in the first two groups were the cause for the insignificant outcome of the Mann-Whitney tests. Moreover, the size of both groups was probably too small. The reasons for their small sizes are mentioned in section 4.1.2. The insignificance and the small size were the reasons that no conclusions were drawn from the results of the tests on these two groups.

The speech therapy students appeared to have been helped by the presence of intonation. Apparently they were able to utilize the extra cues which were given by the intonation of the sentences. In this respect they also compare favourably with the naive listeners of the previous test, who did not distinguish between monotonous and intonated EL speech.

4.6 Summarizing remarks

In tests of the kind described above it would seem advisable to use naive listeners, since they have no special background which could help or interfere with the listening task, be it in a test on naturalness, or on intelligibility of EL speech.

In general terms it may be stated that for naive listeners the advantage of intonated EL speech over monotonous EL speech is especially clear with respect to the naturalness of pitch, rather than to the intelligibility of EL speech. This was somewhat surprising, as it seemed probable that, since EL speech differs qualitatively from laryngeal speech in so many other respects apart from having no intonation, the aspect of the naturalness of pitch does not play such an important role for naive listeners.

The effect of incorrect intonation is even stronger on the intelligibility than on the naturalness of intonated EL speech. In both cases incorrect intonation should be avoided strongly. Training to achieve a high level of proficiency in intonating EL speech is imperative.

The main goal of the introduction of intonation into EL speech, viz. to make it more natural, appears to have been its main improvement: naturalness benefited more from the presence of intonation than intelligibility did. Intelligibility at best benefits indirectly from the presence of intonation: listening to correctly intonated EL speech seems to be less tiring and therefore easier to concentrate on.

[illegible] those which were given [illegible] the sentences. In this respect they also compare favorably with the naive listeners of the previous test, who did not distinguish between monotonous and intonated EL speech.

[illegible] mentioned above it would seem [illegible] listeners, since they have [illegible] background [illegible] [illegible] EL speech.

In general terms it may be stated that for naive listeners the advantage of intonated EL speech over monotonous EL speech is especially [illegible] with respect to the naturalness of pitch [illegible] to the intelligibility of EL speech. This was somewhat surprising, as it seemed probable that, since EL speech differs qualitatively from laryngeal speech in so many other respects apart from having no intonation, the aspect of the naturalness of pitch does not play such an important role for naive listeners.

The effect of incorrect intonation is even stronger on the intelligibility than on the naturalness of intonated EL speech. In both cases incorrect intonation should be avoided strongly. Training to achieve a high level of proficiency in intonating EL speech is imperative.

The main goal of the introduction of intonation into EL speech, viz. to make it more natural, appears to have been its main improvement: naturalness benefited more from the presence of intonation than intelligibility did. Intelligibility at best benefits indirectly from the presence of intonation: listening to correctly intonated EL speech seems to be less tiring and therefore easier to concentrate on.

Chapter 5

CONCLUSIONS AND PROSPECTS

5.1 Introduction

In the first part of this final chapter the findings of this study will be reviewed with respect to the main issue, and to the intonation learning program, with special emphasis on the question whether the Dutch intonation grammar can be taught for practical purposes. Furthermore, the perceptual aspect of intonation in EL speech, and the properties of the EL device will be discussed. In the second part of this chapter some suggestions will be made for possible further development of the EL prototype and the intonation learning program.

5.2 Conclusions

5.2.1 The learning program

The working hypothesis for this study was: limited conscious knowledge of the perceptual and productional aspects of pitch inflection is necessary for adequate use of the pitch inflection control on an EL. A learning program was set up to meet these requirements. As described in section 3.5 the two EL speakers who completed the program did produce speech with correct, and therefore adequate pitch inflection. As such the goal of this study appears to have been reached. The word 'adequate' relates to the role pitch inflection has in the production of pitch accents.

Although explicit rules were provided with which pitch patterns could be made, it was still left to the speaker to decide where to place pitch accents. At present not much is known about the reason why pitch accents are made and where they should be positioned in a sentence. Some attempts to find out about the communicative function of pitch accents are described by Nooteboom, Kruyt & Terken (1981).

In the intonation program EL speakers were taught in an implicit way to rely on their intuitions about accentuation. In figure 3.3 an overview was given of the successive stages leading from reading aloud sentences with given pitch patterns via sentences with pitch accents indicated, to sentences without any such information. For each stage the initial training sentences were the same. Thus the testees were implicitly taught how to identify and produce pitch accents. In fact, they did not have to learn anything new, but rather how to apply the knowledge which was already present in a different way, i.e. they learnt how to produce pitch accents by 'digital' control instead of by the vocal musculature.

It will be obvious that in the approach described, the function of pitch inflection is not regarded in the light of its linguistic aspects, such as sentence meaning or syntactic structure, but rather in a pragmatic sense to the effect that pitch accents serve the purpose of highlighting one or more words in a sentence which for some reason or other deserve to be made prominent.

An assumption made when the intonation program was set up, was that in this particular situation where knowledge was already present but had to be brought under conscious control, a cognitive approach towards this aim would be suitable. The approach was adopted from the Dutch intonation course by Collier & 't Hart (1981). This course and the intonation learning program were therefore based on the same principle. Since what is relevant in the perception of pitch movements is a reflection of what was intended in its production, it was assumed that an explicit description of relevant pitch movements is amenable to learning by productionally handicapped Dutch native speakers. From the results of this study it can be concluded that this is indeed the case.

During its application the various parts of the intonation learning program have often been modified. The relative importance of its components seemed to vary with each testee. E.g. speaker B (cf. section 3.5) could build patterns and draw them once rules to this end had been provided. Speaker C was even less adept in using visual representations of pitch variations, which in his case ruled out the application of visual feedback. With respect to the latter device the conclusion can be drawn that, despite favourable influences it may have, the main and most important feedback for adequate pitch inflection is by ear.

A possible exception to this statement could be found in those cases where hearing loss considerably affects pitch perception, although skill in transcribing pitch contours on the part of the therapist may obviate the need for pitch extraction and display equipment.

The reason why an intonation learning program is at all necessary is clear from the effects found in the perceptual evaluation of intonated EL speech: incorrect intonation had a strong negative influence on perceived naturalness of the speech produced and even more so on its intelligibility. Supplying an EL speaker with a pitch controlled EL without providing necessary instructions would undoubtedly create such a result.

Since this study has the character of a pilot study we did not address the question of how many EL speakers would be able to produce intonated speech with the method used here, for that would imply that this method had already proved to be adequate. We deliberately restricted our investigations to solve the problem whether the method would work at all, or stated a little more properly, whether the intonation description by the Dutch school could be applied in speech rehabilitation.

5.2.2 The semi-automatic EL device

Another outcome of the perceptual evaluation of EL speech concerns the properties of the EL device itself. Both the possibility of producing pitch movements and the automatic declination function proved to be important for the acceptability of speech produced with the device.

When pitch production with this device is compared with the situation in normal speech there is a pronounced similarity. In normal speech, declination control comes about automatically, but declination reset is under voluntary control by the speaker, e.g. when pausing for breath. In speech production with the EL device the same applies. The declination function is automatic and therefore not under control of the speaker. In this respect there is a difference with normal speech where the speaker has some implicit control over the slope of the declination, whereas with the EL it is fixed. Declination reset is controlled by the speaker, it is concomitant with switching the device off and on to make a silent pause. This is not necessary for breathing, for the speaker can keep on talking as long as the device remains switched on. Usually, however, there is a link with breathing which is either a habit remaining from the pre-operative period, or the result of speech therapy to induce natural phrasing.

One might ask whether it really was necessary to develop a new EL. Would not it have been possible to train a laryngectomee to use the pitch inflection control on the Western Electric device? The answer to this question is in the first place that the resulting speech would lack declination and, secondly, that in the absence of automatic control of the pitch movements learners would have to acquire the handling of appropriate pitch excursion, whether falls or rises, with their respective slopes besides learning where to put them in the utterance. Making a pitch movement on the correct syllable was already large enough a burden for the EL speaker; therefore it was decided not to try and place the pitch movement in the optimal position within the syllable.

The semi-automatic pitch control of the EL device was developed on the same principle which was present in the setup of the learning program. It was designed to produce only perceptually relevant pitch movements.

5.3 Suggestions

Some of the suggestions which were the logical consequence of the present study, had already been carried out when this dissertation was written. This especially concerns the development of the semi-automatic EL. Therefore, in the first two sections under this heading some mechanical adjustments and electronic modifications will be described. Consecutively, the further development of the learning program will be discussed. Finally, the question will be addressed whether the approach developed in this study can be applied to other languages as well.

5.3.1 Mechanical adjustments

As it was, the EL device served its purpose very well: the testees were able to produce intonated speech with it. But its design was still far from optimal. Some points which needed to be taken care of were its size, which was about 21 cm, while the Servox EL measures about 11 cm. The pitch control button extended too far from the body of the device. The switches used were fairly large, thus the push-and-slide construction of the pitch control button was too space-consuming and awkward (although it never broke down). Possibly the electronic circuit could also be improved as to its size and as to power consumption.

Consequently a request to subsidize a project to optimize the semi-automatic EL was submitted to the Dutch National Cancer Fund, which was approved. This project lasted 14 months and the outcome was a prototype which was similar to the Servox EL, of which the vibrating head had been utilized with all the prototypes tested. On the basis of some informal testing the Servox vibrator head proved to be very efficient as regards energy consumption. Trying to improve it with regard to the sound it produced was considered too specialized a task to be included in the project. The question of the efficacy of the vibrating head as an artificial sound source simulating vocal cord characteristics constitutes a research topic in its own right.

A major difference with the original EL prototype was that the mechanical construction of the pitch control button was changed. It proved to be unreliable if small components were utilized. A test with a few working models with different pitch control buttons showed that there was no preference for a single button, which combined two functions, over a construction with two buttons. The latter construction was mechanically more straightforward and therefore more reliable. Thus the final prototype had one pushbutton for the on/off function which was combined with the silent declination reset, and one pushbutton for effecting pitch movements by switching from one declination level to the other and vice versa.

5.3.2 Electronic modifications

The design of the electronic circuit was changed to one with fewer components and with independent preset controls for adjustment of overall pitch range, output volume, slope of the declination levels, and the distance between them. Power consumption was also reduced as far as possible. One wish remained: to further reduce the size of the electronics into an integrated circuit. To this end two students at the Electronics-Systems-Automation-Technology laboratory (ESAT) of the university of Louvain (Belgium) are occupied at present in designing a layout for such an IC.

5.3.3 The learning program

As regards the learning program it will be clear that it still needs a lot of improvement. In section 5.2.1 it was mentioned that the relative importance of its various parts varied per testee. This calls for a very flexible setup of the program to allow for the laryngectomee's capability of abstraction, possible hearing deficiencies, skill in dealing with language, etc. In general, the program could gain from considering therapeutical aspects, as it had been set up with only phonetic-didactic principles in mind. A related wish concerns the persons who are to train the EL speakers. This is a speech therapist's task, and rather than a phonetician's. So in order to reach a larger number of EL speakers it is desirable to instruct a group of therapists first. These wishes have culminated into a successful application for a research grant, again directed to the Dutch National Cancer Fund, in which the author and a speech therapist are to develop a pre-final learning program which meets the above requirements, and to train a group of therapists and support them during the application of the program. This should then lead to a final intonation learning program and a coursebook for EL speakers.

5.3.4 Use in other languages

In this discussion one question has not been addressed yet. The semi-automatic EL is built with specifications related to Dutch intonation. What are the prospects for applying the device for other languages? Declination can be considered perceptually relevant in almost any language. Conceivably, fitting a Western Electric EL nr. 5 A or B with an automatic declination function might already improve the intonational acceptability of speech produced with the aid of such an adapted EL (Van Geel, 1982). Nevertheless, the use of discrete pitch movements in EL speech is to be preferred to a continuous pitch inflection control for reasons given in section 2.8. A necessary condition for application of the concept of discrete pitch movements is that, for the language concerned, there should be an intonation grammar built along the lines of the Dutch school. Once that is available a choice for a subset of pitch movements can be made on the basis of frequency of occurrence data. It would seem that only a subset can be used to keep the pitch inflection control simple and easy to use. Thus it might be possible that a semi-automatic EL for British English will be developed as a consequence of research currently performed in the tradition of the Dutch school.

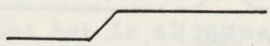
However, since the intonation of British English appears to be based on three declination levels instead of two for the Dutch language, the pitch control circuit would be more complex by necessity. Hopefully the incorporation of a canonical pitch pattern will allow a design of a simple pitch control circuit in an EL for British English intonation.

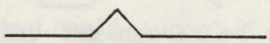
5.4 Final words

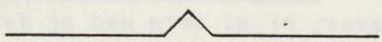
It is hoped that the efforts to introduce intonation in EL speech may result in a type of speech which sounds less robot-like and that this may contribute to a more widespread acceptance of EL speech.

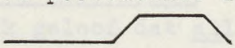
Appendix A

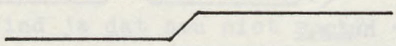
SENTENCES USED FOR TRAINING IN READING ALOUD (cf. 3.1.3)

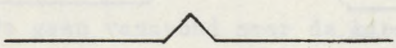
1. 

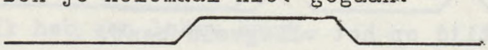
2. 
Ga je mee vanavond?

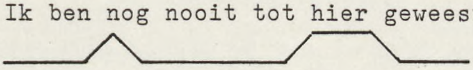
3. 
Je portemonnee is gevallen.

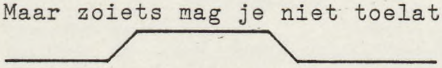
4. 
Dat is heel mooi.

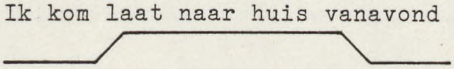
5. 

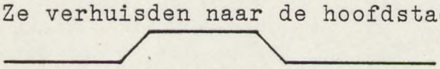
6. 
Ben je helemaal niet gegaan?

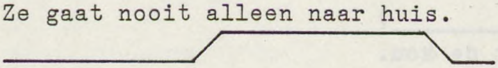
7. 
Ik ben nog nooit tot hier geweest.

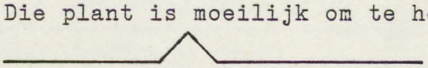
8. 
Maar zoiets mag je niet toelaten.

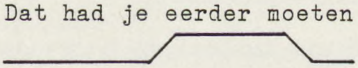
9. 
Ik kom laat naar huis vanavond

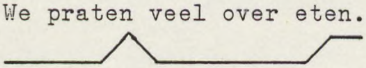
10. 
Ze verhuisden naar de hoofdstad.

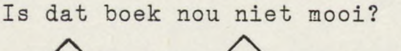
11. 
Ze gaat nooit alleen naar huis.

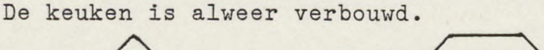
12. 
Die plant is moeilijk om te houden.

13. 
Dat had je eerder moeten doen.

14. 
We praten veel over eten.

15. 
Is dat boek nou niet mooi?

16. 
De keuken is alweer verbouwd.

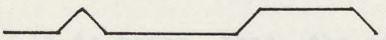
17. 
De telefoon gaat vandaag wel heel vaak.

Appendix B

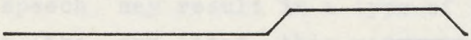
PERCEPTION AND PRODUCTION TESTS OF THE LEARNING PROGRAM (cf. 3.1.4)

(material taken from Collier & 't Hart, 1981)

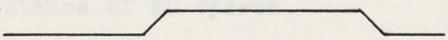
<u>Perception test</u> (with pitch patterns to be recognized between parentheses):

()

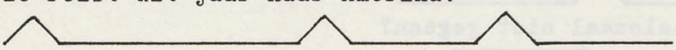
1. Het eten is nog niet klaar.

()

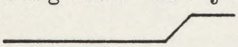
2. Hij komt vanavond laat naar huis.

()

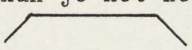
3. Ze reist dit jaar naar Amerika.

()

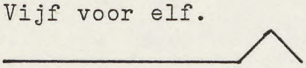
4. Zorg maar dat je op tijd op het vliegveld bent.

()

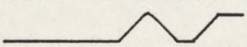
5. Kun je het lezen?

()

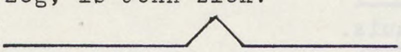
6. Vijf voor elf.

()

7. Daar heb ik hem nooit.

()

8. Zeg, is John ziek?

()

9. Het valt wel mee met de kou.

Production test (first presentation with syllables to be accented underlined; second presentation with pitch patterns given):

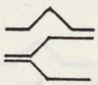
1. Dat heb ik al duizend keer gezegd.
2. Nee beslist niet.
3. Heb je hem niet in je jaszak?
4. Ik geloof dat alles in orde is.
5. Vind je dat nou niet mooi?
6. We gaan vanavond naar de bioscoop.
7. Ik heb een lekke band.

Appendix C

RULES FOR PRODUCING PITCH PATTERNS (cf. 3.5.2)

1. Main principle: rises and falls alternate.
2. A pattern is determined by the number of accents in an utterance and by the main principle.
3. A pattern mostly begins and ends at the low declination level, thus the first pitch movement mostly is a rise (in 97% of all patterns), and the last one mostly a fall (92%).
4. An accent can be made by three alternative realizations, depending on the main principle:

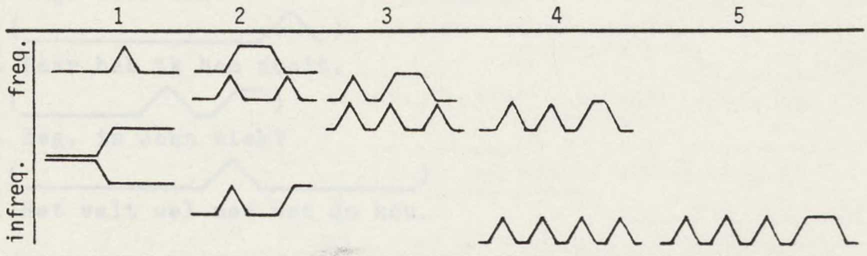
a)
b)
c)



5. There are also non-accent lending pitch movements:
- any fall which is not the final pitch movement in an utterance is not accent lending, e.g. in precursive accent peaks, or at sentence boundaries, or as delayed falls.
- a rise at the end of a pattern, e.g. a continuation rise, or at the end of many question sentences.

MOST FREQUENT PATTERNS FOR A GIVEN NUMBER OF ACCENTS

number of accents:



The height along the Y-axis indicates the relative frequency of occurrence of each pattern.

Appendix D

SENTENCES USED IN THE LISTENING TEST OF 4.1.2 WITH THE ACCOMPANYING PITCH PATTERNS.

Versions P, C and S

Ja, dat wordt het zeker.

Het gaat langzaam vooruit.

We gaan gewoon zo door.

Nu gaat het nog een beetje krampachtig, maar, ik stel me voor,

dat, na verloop van tijd dat wel weer beter zal gaan.

Het lezen gaat bij mij weer beter dan voorheen.

Voor het lezen heb ik hem nu niet meer nodig.

Dan hebben we niet voor niets geoefend.

Version I

Ja, dat wordt het zeker.

Het gaat langzaam vooruit.

We gaan gewoon zo door.

Nu gaat het nog een beetje krampachtig, maar, ik stel me voor,

dat, na verloop van tijd dat wel weer beter zal gaan.

Het lezen gaat bij mij weer beter dan voorheen.

Voor het lezen heb ik hem nu niet meer nodig.

Dan hebben we niet voor niets geoefend.

Appendix E

SENTENCES USED IN THE LISTENING TEST OF 4.2.2

Ja, dat wordt het zeker.

Het gaat langzaam vooruit.

Voor het lezen heb ik hem nu niet meer nodig.

Dan hebben we niet voor niets geoefend.

Daar is zeer veel belangstelling voor.

Hoe vind je dat nou?

De trein naar Utrecht gaat elk halfuur.

Het is waar.

Ik begin er steeds meer plezier in te krijgen.

Ik vind het net of de zaal nog niet af is.

Appendix F

SENTENCES USED IN THE LISTENING TEST OF 4.4.2

Ja, dat wordt het zeker.

Het gaat langzaam vooruit.

We gaan gewoon zo door.

Het lezen gaat bij mij weer beter dan voorheen.

Voor het lezen heb ik hem nu niet meer nodig.

Dan hebben we niet voor niets geoefend.

Daar wen je wel aan vind ik tenminste.

Dit doe ik nog een keer.

Er is vanmiddag een hoop verkeer op de weg.

Daar is zeer veel belangstelling voor.

Appendix G

SENTENCES USED IN THE LISTENING TEST OF 4.5.2

(each second sentence was presented in a monotonous version)

Ja, dat wordt het zeker.
Ja, het gaat weer beter.

Het gaat langzaam vooruit.
Hij komt altijd te laat.

Voor het lezen heb ik hem nu niet meer nodig.
En tenslotte moest hij nog de deur dichtdoen.

Dan hebben we niet voor niets geoefend.
Toen hebben we fijn aan zee gezeten.

Daar is zeer veel belangstelling voor.
Toen brak toch nog het zonnetje door.

Hoe vind je dat nou?
Wat dacht je ervan?

De trein naar Utrecht gaat elk halfuur.
Een gele auto valt heel goed op.

Het is waar.
Ik ben klaar.

Ik begin er steeds meer plezier in te krijgen.
Er gebeurt vandaag verder niets bij mijn weten.

Ik vind het net of de zaal nog niet af is.
Hij vroeg meteen of de auto al klaar was.

Summary

This thesis deals with the problem which faces someone who has lost his vocal cords, called a laryngectomee, and who utilizes an electronic artificial larynx (EL): the speech he produces is robot like, i.e. it is monotonous. Although there exists an EL device which permits pitch variations to be produced, this feature is rarely, if ever, used. It was assumed that lack of conscious knowledge of pitch phenomena in speech caused this problem. Therefore it was decided to try and develop an intonation learning program for EL speakers which aims at remedying the monotony problem by providing some intonation theory to enable them to use the pitch inflection control of the EL effectively. This program is based on Dutch intonation research. Therefore, in chapter one an introduction into Dutch intonation theory is presented.

The second chapter also has an introductory function. It serves to acquaint the uninitiated reader with speech rehabilitation possibilities for the laryngectomee. It deals with various ways of voice production and especially with the properties of electrolaryngeal speech. Also the design of an EL prototype is described featuring a semi-automatic pitch control which leaves only timing as a parameter to be controlled by the user. This is taken care of in the learning program. At the end of the chapter an outline for the present study is given.

In the third chapter an intonation learning program is described. First its development is reported and then its application on non-laryngectomees by way of a pilot test. Finally, after some necessary revisions, its application on three EL speakers is described. This has resulted in further revisions some of which were made during the actual application of the program. Two of the three speakers mastered spontaneous and correctly intonated EL speech; the third had to discontinue the program due to illness.

In chapter four listening tests are described which were set up to assess the merits of intonation in EL speech as regards its naturalness and its intelligibility. The tests also

served to evaluate the design of the EL prototype. The basic conclusion was that only EL speech with correct pitch patterns was judged as having a natural intonation when compared to, among others, monotonous EL speech. Incorrect intonation, e.g. because of inaccurate timing in making pitch movements, was considered just as unnatural as no intonation, which stresses the importance of adequate training in the production of correctly intonated EL speech. The effect of incorrect intonation on the intelligibility of this type of speech was even stronger: it was distinctly less intelligible than monotonous or correctly intonated EL speech.

In the final chapter it is concluded that it is possible to teach an EL speaker to produce correct pitch inflection, based on Dutch intonation research. Some suggestions are made for improving the intonation learning program and the EL prototype.

Samenvatting

Dit proefschrift houdt zich bezig met de problemen van iemand die niet meer beschikt over zijn stembanden, een gelaryngectomeerde ofwel stembandloze, en die gebruik maakt van een electronische kunstlarynx (EL): de daarmee geproduceerde spraak klinkt robotachtig, d.w.z. is monotoon. Hoewel er een EL bestaat waarmee het mogelijk is om toonhoogtevariaties te produceren, wordt er zelden of nooit van deze mogelijkheid gebruik gemaakt. Verondersteld werd dat een gebrek aan bewuste kennis van toonhoogteverschijnselen in de spraak de oorzaak was voor dit probleem. Daarom werd besloten om te proberen een intonatieleerprogramma te ontwikkelen voor EL sprekers, dat als doel heeft het probleem van de monotonie aan te pakken met behulp van intonatietheorie om hen zo in staat te stellen op effectieve wijze de toonhoogteregeling van de EL te bedienen. Dit programma is gebaseerd op intonatie onderzoek van het Nederlands. Daarom bestaat hoofdstuk een uit een introduktie in de Nederlandse intonatie theorie.

Het tweede hoofdstuk is eveneens een introduktie. Daarin wordt de lezer bekend gemaakt met de verschillende mogelijkheden van spraakrehabilitatie die er zijn voor stembandlozen. Het gaat in op verschillende manieren van spraakproduktie, waarvan EL spraak uitgebreid wordt toegelicht. Ook wordt een ontwerp voor een EL prototype beschreven dat gekenmerkt wordt door een halfautomatische toonhoogtesturing, waarbij het enige wat aan de gebruiker nog overgelaten wordt, is, dat de juiste momenten voor het maken van toonhoogtebewegingen door hemzelf moeten worden bepaald. Hierin voorziet het leerprogramma. Aan het eind van dat hoofdstuk wordt de opzet van het onderzoek beschreven.

Het derde hoofdstuk is gewijd aan een intonatieleerprogramma. Eerst de ontwikkeling daarvan, dan de toepassing op enkele niet-stembandlozen bij wijze van eerste verkenning en tenslotte, nadat enkele noodzakelijke wijzigingen waren aangebracht, de toepassing op een drietal EL sprekers. Als gevolg hiervan leken opnieuw enkele wijzigingen nodig, waarvan een

gedeelte al werd aangebracht tijdens de toepassing zelf. Twee van de drie sprekers leerden op spontane wijze correct intonerend te spreken. De derde moest wegens ziekte voortijdig het programma beëindigen.

In het vierde hoofdstuk worden luisterproeven beschreven waarmee de waarde van intonatie voor EL spraak werd beoordeeld met betrekking tot de natuurlijkheid en de verstaanbaarheid ervan. Tevens werd het ontwerp van het EL prototype geëvalueerd. De hoofdconclusie was dat EL spraak met correcte intonatie inderdaad als natuurlijk werd beschouwd, wanneer het vergeleken werd met o.a. monotone EL spraak. Incorrect geïntoneerde EL spraak, bijvoorbeeld als gevolg van slechte timing van de toonhoogtebewegingen, werd net zo onnatuurlijk gevonden als monotone spraak, hetgeen het belang van een adekwate training voor het produceren van correct geïntoneerde EL spraak alleen maar onderstreept. Het effekt van incorrecte intonatie op de verstaanbaarheid van EL spraak was nog sterker: aldus geïntoneerde spraak bleek beduidend minder goed verstaanbaar dan monotone en correct geïntoneerde spraak. Het onderliggende idee voor het ontwerp van het EL prototype bleek aan de verwachtingen te voldoen.

In het laatste hoofdstuk wordt geconcludeerd dat het inderdaad mogelijk is om een EL spreker te leren correct geïntoneerde spraak te produceren, uitgaande van Nederlands intonatieonderzoek. Ook worden enkele suggesties gedaan ter verbetering van het leerprogramma en het EL prototype.

REFERENCES

ABBERTON, E. & FOURCIN, A.J. (1975)
Visual feedback and the acquisition of intonation, in: Lenneberg & Lenneberg (1975): 157-165.

BARNEY, H.L., HAWORTH, F.E. & DUNN, H.K. (1959)
An experimental transistorized artificial larynx, Bell System Technical Journal 38: 1337-1356.

BENNETT, S. & WEINBERG, B. (1973)
Acceptability ratings of normal, esophageal, and artificial larynx speech, JSHR 16: 608-615.

VAN DEN BERG, J., MOOLENAAR-BIJL, A.J. & DAMSTE, P.H. (1958)
Oesophageal speech, Folia Phoniatrica 10: 65-84.

BLOM, E.D. & SINGER, M.I. (1979)
Surgical-prosthetic approaches for postlaryngectomy voice restoration, in: Keith & Darley (1979): 251-276.

BOLINGER, D.L. (1958)
A theory of pitch accent in English, Word 14: 109-149.

BRANDERUD, P. GALYAS, K. & SVENSSON, S.G. (1974)
Experimental device for control of intonation with an artificial larynx, STL-QPSR 4: 38-41.

BROWN, G., CURRY, K.L. & KENWORTHY, J. (1980)
Questions of intonation, (Croom Helm, London).

BURGER, H. & KAISER, L. (1925)
Speech without a larynx, Acta Otolaryngol. 8: 90-116.

COHEN, A. & 't HART, J. (1965)
Perceptual analysis of intonation patterns, Proc. Vth Int. Congr. Acoust., Liège.

COHEN, A. & 't HART, J. (1967)
On the anatomy of intonation, Lingua 19, 177-192.

COHEN, A., COLLIER, R. & 't HART, J (1982)
Declination: construct or intrinsic feature of speech pitch?, Phonetica 39: 254-273.

COHEN, A., Van den BROECKE, M.P.R. & Van GEEL, R.C. (to appear)
A study of pitch phenomena and applications in electrolarynx speech, in: Lass (to appear).

COLLIER, R. (1972)
From pitch to intonation, (unpubl. diss., Louvain University).

COLLIER, R. (1975)
Physiological correlates of intonation, JASA 58: 249-255.

COLLIER, R. & 't HART, J. (1981)
Cursus Nederlandse Intonatie, (Acco, Louvain, Belgium).

CRYSTAL, D. (1969)
Prosodic systems and intonation in English, (Cambridge University Press).

CURRY, E.T, SNIDECOR, J.C. & ISSHIKI, N. (1973)
Fundamental frequency charateristics of Japanese Asai speakers, Laryngoscope 83: 1759-1763.

DAMSTE, P.H. (1958)
Oesophageal speech after laryngectomy, (Gebr. Hoitsema, Groningen).

DAMSTE, P.H. (1975)
Methods of restoring the voice after laryngectomy, Laryngoscope 85: 649-655.

DAMSTE, P.H. & HOSANG-JACOBSON, A. (1978)
Toepassing van de kunstmatige larynx onmiddellijk na laryngectomie, Logopedie en Foniatrie 50: 265-270.

DENES, P.B. (1959)
A preliminary investigation of certain aspects of intonation, Language and Speech 2: 106-122.

DENES, P.B. & PINSON, E.N. (1973)
The Speech Chain, (Anchor Press/Doubleday, New York).

DIEDRICH, W.M. (1968)
The mechanism of esophageal speech, Annals of the New York Academy of Sciences 155: 303-317.

DIEDRICH, W.M. & YOUNGSTROM, K.A. (1966)
Alaryngeal speech, (Ch. C. Thomas, Springfield, Ill.).

DONEGAN, J.O., GLUCKMAN, J.L. & SINGH, J. (1981)
Limitations of the Blom-Singer technique for voice restoration, Ann. Otol. Rhinol. Laryngol. 90: 495-497.

EDWARDS, A.L. (1957)
Techniques of attitude scale construction, (Appleton- Century-Crofts, New York).

EVERETT, R.L. & BAILEY, B.J. (1982)
Implanted electrolaryngeal prosthesis, in: Sekey (1982): 215-235.

FIRESTONE, F.A. (1940)
An artificial larynx for speaking and choral singing by one person, JASA 11: 357-361.

FRY, D.B. (1958)
Experiments in the perception of stress, Language & Speech 1: 126-152.

GALYAS, K., BRANDERUD, P. & McALLISTER, R. (1982)
The 'Intonator': development of an electrolarynx with intonation control, in: Sekey (1982): 184-189.

GANDOUR, J. & WEINBERG, B. (1982)
Perception of contrastive stress in alaryngeal speech, J. of Phon. 10: 347-359.

GARDNER, W.H. (1971)
Laryngectomee speech and rehabilitation, (Ch.C. Thomas, Springfield, Ill.).

GATES, G.A. & HEARNE, E.M. (1982)
Predicting esophageal speech, Ann. Otol. Rhinol. Laryngol. 91: 454-457.

Van GEEL, R.C. (1980)
Vowel production with two types of electrolarynges - an acoustic and perceptual comparison, Progr. Rep. Inst. of Phon. Utrecht 5(1): 3-9.

Van GEEL, R.C. (1982)
Semi-automatic pitch control for an electrolarynx, in Sekey (1982): 190-197.

Van GEEL, R.C. & Van der WERFF, M.H. (1980)
An experimental design of a pitch control for an electrolarynx, Progr. Rep. Inst. of Phon. Utrecht 5(1): 60-63.

GREEN, G. & HULTS, M. (1982)
Preferences for three types of alaryngeal speech, JSHD 47: 141-145.

HALLIDAY, M.A.K. (1970)
A course in spoken English: intonation, (Oxford University Press).

't HART, J. (1966)
Perceptual analysis of Dutch intonation features, IPO Ann. Progr. Rep. 1: 47-51.

't HART, J. (1979)
Explorations in automatic stylization of F0 curves, IPO Ann. Progr. Rep. 14: 61-65.

't HART, J. (1981)
Een cognitieve methode voor het leren van Nederlandse intonatie, Toegepaste Taalwetenschap in Artikelen 9: 98-111.

't HART, J. & COHEN, A. (1973)
Intonation by rule: a perceptual quest, J. of Phon. 1: 309-327.

't HART, J. & COLLIER, R. (1975)
Integrating different levels of intonation analysis, J. of Phon. 3: 235-255.

't HART, J. & COLLIER, R. (1978)
A course in Dutch intonation, IPO Ann. Progr. Rep. 13: 31-35.

IPA (1949)
The principles of the International Phonetic Association, (Dept. of Phonetics, University College, London).

ISSHIKI, N. & TANABE, M (1972)
Acoustic and aerodynamic study of a superior electrolarynx speaker, Folia Phoniatrica 24: 65-76.

KATZ, P. (1982)
(reference in IAL News 27(2)).

KEITH, R.L. & DARLEY, F.L. (1979)
Laryngectomy rehabilitation, (College Hill Press, Houston).

KLATT, D.H. (1973)
Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception, JASA 53: 8-16.

LASS, N. (to appear)
Speech and language: advances in basic research and practice, (Academic Press, New York).

LAUDER, E. (1968)
The laryngectomee and the artificial larynx, JSHD 33: 147-157.

LEBRUN, Y. (1973)
The artificial larynx, (Swets & Zeitlinger BV, Amsterdam).

LEHISTE, I. (1970)
Suprasegmentals, (MIT Press, Cambridge, Mass.).

LENNEBERG, E.H. & LENNEBERG, E. (1975)
Foundations of language development, vol. 2, (Academic Press, New York).

LIEBERMAN, P. (1965)
On the acoustic basis of the perception of intonation by linguists, Word 21: 40-54.

LOWRY, L.D. (1981)
Aritificial larynges: a review and development of a prototype self-contained intra-oral artificial larynx, Laryngoscope 91: 1332-1355.

MAEDA, S. (1976)
A characterisation of American English intonation, Ph.D. thesis (MIT , Cambridge, Mass.).

MALBECK, E. & SCHOSSHAUER, B. (1960)
Phoniatrische Nachuntersuchungen chordektomierter Patienten, HNO 8: 201-205.

McCROSKEY, R.L. & MULLIGAN, M. (1963)
The relative intelligibility of esophageal speech and artificial larynx speech, JSHD 28: 37-41.

MOOLENAAR-BIJL, A. (1953)
Connection between consonant articulation and the intake of air in oesophageal speech, Folia Phoniatrica 5: 212-216.

MORTON, J. & JASSEM, W. (1965)
Acoustic correlates of stress, Language & Speech 8: 159-181.

MILLER, A.H. (1962)
First experiences with the Asai technique for vocal rehabilitation after total laryngectomy, in: Snidecor (1962): 50-57.

NIJDAM, H.F. ANNYAS, A.A., SCHUTTE, H.K. & LEEVER, H. (1982)
A new prosthesis for voice rehabilitation after laryngectomy, Acta Otorhinolaryngol. 237: 27-33.

NOOTEBOOM, S. KRUYT, T, & TERKEN, J. (1981)
What speakers and listeners do with pitch accents: some explorations, in: Tretheim (1981): 9-32.

O'CONNOR, J.D & ARNOLD, G.F. (1973)
Intonation of colloquial English, (Longman's, London).

PANJE, W.R. (1981)
Prosthetic vocal rehabilitation following laryngectomy: the voice button, Ann. Otol. 90: 116-120.

PANJE, W.R., VAN DEMARK, D. & McCABE, B.F. (1981)
Voice button prosthesis rehabilitation of the laryngectomee: additional notes, Ann. Otol. Rhinol. Laryngol. 90: 503-505.

PERRY, A., CHEESMAN, A.D., EDEN, R. (1982)
A modification of the Blom Singer valve for restoration of voice after laryngectomy, J. of Laryngol. and Otol. 96: 1005-1111.

De PIJPER, J.R. (1979)
Close copy stylizations of British English intonation contours, IPO Ann. Progr. Rep. 14: 66-71.

De PIJPER, J.R. (to appear)
Modelling British English Intonation: a bottom up approach, diss., (Utrecht University).

PIKE, K.L. (1945)
The intonation of American English, (University of Michigan Press, Ann Harbor).

RIESZ, R.R. (1930)
Description and demonstration of an artificial larynx, JASA 1: 273-279.

ROTHMAN, H.B. (1978a)
Analysing artificial electronic larynx speech, in Salmon & Goldstein (1978): 87-111.

ROTHMAN, H.B. (1978b)
Approaches to treatment, part E, in: Salmon & Goldstein (1978): 131-137.

SALMON, S.J. (1978)
Approaches to treatment, part F, in: Salmon & Goldstein (1978): 137-144.

SALMON, S.J. & GOLDSTEIN, L.P. (1978)
The artificial larynx handbook, (Grune & Stratton, New York).

SCHAEFER, S.D. & JOHNS, D.F. (1982)
Attaining functional esophageal speech, Arch. Otolaryngol. 108: 647-649.

SEKEY, A. (1982)
Electroacoustic analysis and enhancement of alaryngeal speech, (Ch. C. Thomas, Springfield, Ill.).

SEKEY, A. & HANSON, R. (1982)
Laryngectomee speech support system with prosodic control, in: Sekey (1982): 166-183.

SHAMES, G.H. & FONT, J. & MATTHEWS, J. (1963)
Factors related to speech proficiency of the laryngectomized, JSHD 28: 273-287.

SHAPIRO, M.J. & RAMANATHAN, V.R. (1982)
Trachea stoma vent voice prosthesis, Laryngoscope 92: 1126-1129.

SINGER, M.I., BLOM, E.D. & HAMAKER, R.C. (1981)
Further experience with voice restoration after total laryngectomy, Ann. Otol. Rhinol. Laryngol. 90: 499-502.

SISSON, G.A., McCONNEL, F.M.S., LOGEMANN, J.A. & YEH, S. (1975)
Voice rehabiliation after laryngectomy: results with the use of a hypopharyngeal prosthesis, Arch. Otol. 101: 178-181.

SMITH, S. (1962)
On artificial voice production, Proc. IVth Int. Congr. Phon., (Helsinki): 96-110.

SNIDECOR, J.C. (1962)
Speech rehabilitation of the laryngectomized, (Ch. C. Thomas, Springfield, Ill.).

TAIT, V. & TAIT, R.V. (1959)
Speech rehabilitation with the oral vibrator, Speech Pathology and Therapy 2: 64-69.

TAUB, S. (1975)
Air bypass voice prosthesis for vocal rehabilitation of laryngectomees, Ann. Otol. Rhinol. Laryngol. 84: 45-48.

TRAGER, G.L. & SMITH, H.L. (1951)
An outline of English structure, Studies in Linguistics Occasional Papers 3.

TRETHEIM, T. (1981)
Nordic prosody: papers from a symposium, (Tapir, Trondheim).

VAN RIPER, Ch. (1972)
Speech correction: principles and methods, (Prentice Hall, Englewood Cliffs, New Jersey).

VEGA, M.F. (1975)
Larynx reconstructive surgery - a study of three years findings - a modified surgical technique, Laryngoscope 85: 866-881.

WEINBERG, B. (1980)
Readings in speech following laryngectomy, (University Park Press, Baltimore, Maryland).

WEINBERG, B. & BOSMA, J.F. (1970)
Similarities between glossopharyngeal breathing and injection methods of air intake for esophageal speech, JSHD 35: 25-32.

WEINBERG, B. & WESTERHOUSE, J. (1971)
A study of buccal speech, JSHR 14: 652-658.

WEINBERG, B. & WESTERHOUSE, J. (1973)
A study of pharyngeal speech, JSHD 38: 111-118.

WEISS, M.S., YENI-KOMSHIAN, G.H., HEINZ, J.M. (1979)
Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx, JASA 65: 1298-1308.

WELLS, R. (1945)
The pitch phonemes of English, Language 21: 27-39.

WILLEMS, L.F. (1966)
The Intonator, IPO Ann. Progr. Rep. 1: 123-125.

WILLEMS, N.J. (1982)
English intonation from a Dutch point of view, (Foris Publications, Dordrecht).

WINER, B.J. (1971)
Statistical principles in experimental design, 2nd ed., (McGraw- Hill, New York).

YOUNG, I.M. & HARBERT, F. (1970)
Noise effect on speech discrimination scores, J. Audit. Res. 10: 127-131.

ZWITMAN, D. & KNORR, S. (1982)
Intraoral electrolarynx, abstract in: Sekey (1982): 315-316.

VAN RIPER, CH. (1972)
Speech correction: principles and methods, (Prentice-Hall, Englewood Cliffs, New Jersey).

VEGA, M.F. (1975)
Larynx reconstructive surgery - a study of three-year findings - a modified surgical technique, Laryngoscope 85: 866-881.

WEINBERG, B. (1980)
Readings in speech following laryngectomy, (University Park Press, Baltimore, Maryland).

WEINBERG, B. & BOSMA, J.F. (1970)
Similarities between glossopharyngeal breathing and injection methods of air intake for esophageal speech, JSHD 35: 25-32.

WEINBERG, B. & WESTERHOUSE, J. (1971)
A study of buccal speech, JSHR 14: 652-658.

WEINBERG, B. & WESTERHOUSE, J. (1973)
A study of pharyngeal speech, JSHD 38: 111-118.

WEISS, M.S., YENI-KOMSHIAN, G.H., HEINZ, J.M. (1979)
Acoustical and perceptual characteristics of speech produced with an electronic artificial larynx, JASA 65: 1298-1308.

WELLS, R. (1945)
The pitch phonemes of English, Language 21: 27-39.

WILLEMS, L.F. (1966)
The intonator, IPO Annual Progress Report 1: [illegible]

WILLEMS, N.J. (1982)
English intonation from a Dutch point of view, (Foris Publications, Dordrecht).

WINER, B.J. (1971)
Statistical principles in experimental design, 2nd ed., (McGraw-Hill, New York).

YOUNG, I.M. & HARBERT, F. (1970)
Noise effect on speech discrimination score, J. Audit. Res. 10: 127-131.

ZWITMAN, D. & KNORR, S. (1982)
Intraoral electrolarynx, abstract in: Sekey (1982): 315-316.

ZC 2859

Curriculum vitae

15 juni 1953	: geboren te Hilversum
19 mei 1972	: eindexamen Gymnasium-b
9 juni 1978	: doctoraal examen Fonetiek, met bijvakken Foniatrie en Toegepaste Taalkunde
1 april 1979 tot 1 maart 1982	: verbonden aan het Fonetisch Instituut der R.U. Utrecht op het projekt 'Produktie en perceptie van intonatie in kunstlarynx-spraak', gesubsidieerd door ZWO onder nummer 17-21-16
6 september 1982 tot heden	: docent Fonetiek bij de Opleiding Logopedie en Akoepedie te Utrecht
1 april 1983 tot heden	: verbonden aan het Fonetisch Instituut op het projekt 'Toonhoogteleerprogramma voor electrolarynxsprekers', gesubsidieerd door het Kon. Wilhelmina Fonds onder nummer UUKC 82-16

STELLINGEN

1. Oefenen op het eerder of later resp. in- of uitschakelen van een electrolarynx om zo stemloze begin- resp. eindconsonanten te kunnen maken (gesimuleerde "voice onset lag") is een miskenning van het distinktief kenmerk "gespannen", dat evenzeer een onderscheid aanbrengt tussen de zg. "stemhebbende" consonanten als het kenmerk "stem" dat doet.

2. Een vergelijking van de toonhoogteomvang van electrolarynxspraak met die van normale, of slokdarmspraak is pas zinvol, wanneer er in electrolarynxspraak adekwate toonhoogtevariaties gerealiseerd worden (met een daartoe ingericht apparaat); tot dat moment worden er bij electrolarynxspraak hoofdzakelijk artefakten gemeten.

3. Er zijn in het Nederlands voor het leren intoneren in electrolarynxspraak geen expliciete instrukties nodig omtrent het plaatsen van toonhoogteaccenten; men kan op reeds aanwezige intuïties daaromtrent vertrouwen.

4. De wens dat een electrolarynspreker in staat moet zijn om ook "stemloze" spraakklanken te maken is in het licht van het gevolg van de laryngectomie bijna strijdig te noemen met spraakrevalidatie.

5. De uitdrukking "c'est le ton qui fait la musique" geeft goed aan wat de invloed is van het Nederlandse intonatieonderzoek op electrolarynxspraak.

6. Maar al te vaak worden invaliden met een nieuw ontwikkeld hulpmiddel zonder adekwate training "in het diepe gegooid". Dit is een verspilling van tijd en inspanning, zowel van de kant van de ontwikkelaars als van de kant van de teleurgestelde gebruikers.

7. De woordkeus in symptoombeschrijvingen bij patiënten in diverse publikaties is vaak nodeloos indirekt, bv. Everett & Bailey (1982:200): "...caused swallowing difficulty which led to substantial weight loss in the patient."

8. Op zich betrouwbare metingen van aspekten van luister- en spreekvaardigheid in een vreemde taal worden vaak ten onrechte gepresenteerd als valide globale toetsen voor het meten van die taalvaardigheid.

9. Het is eigenlijk verbazingwekkend, gezien de complexiteit van het proces van spreken en verstaan, dat er nog zo veel begrepen wordt van wat er gesproken wordt.

10. Het nut van de posters van Veilig Verkeer Nederland over goede achterverlichting van fietsen, die geplaatst zijn langs voor fietsers niet toegankelijke wegen, is op zijn best twijfelachtig te noemen.

11. Menig nieuw kantoorgebouw blijkt pas bij nadere beschouwing er zo uit te zien alsof het op een maquette heel aardig toonde; men was alleen vergeten het gebouw te bekijken vanuit het perspectief van de mensen op die maquette.

12. Het is merkwaardig dat de laatste jaren bij collectes steeds meer klein geld wordt geschonken, men zou moeten beseffen dat men door toedoen van de geldontwaarding toch al minder geeft.

13. Van de niet op een proefschrift betrekking hebbende stellingen zijn er vele die nauwelijks waarde hebben en slechts voortkomen uit de verplichting tot die stellingen. Daar is deze stelling er een van.

Stellingen behorende bij het proefschrift "Pitch inflection in electrolaryngeal speech", R.C. van Geel, 1 juli 1983.