Vowel duration and a model of lip mandible coordination

Lindblom, B.

journal: STL-QPSR
volume: 8
number: 4
year: 1967
pages: 001-029

http://www.speech.kth.se/qpsr
I. SPEECH PRODUCTION

A. VOWEL DURATION AND A MODEL OF LIP MANDIBLE COORDINATION

B. Lindblom

A classical problem in phonetics is that of vowel duration. The problem arises because of the systematic variations that the acoustic duration of a vowel is known to exhibit as a function of the feature composition of both the vowel itself and the adjacent consonantal environment. For example, everything else being equal,

1. open vowels tend to be longer than close vowels;
2. a vowel is generally longer before a voiced consonant than before a voiceless consonant; and
3. it is longer also before a fricative than before a stop (1, 2).

At present the causes underlying the differences listed under (1) - (3) are largely unknown. Nor can anything be said with certainty about their cross-lingual generality. Consequently phonetic theory has yet to be developed to the point where it predicts and explains the acoustic facts of vowel duration and provides a rationale for regarding them as linguistically determined or as consequences of universal physiological conditions on human speech production. In so far as such an undertaking is successful it is likely to deepen our general understanding of the motor organization of the syllable and to bear on the question of determining the form of phonetic rules in linguistic description.

In the following study we attempt to analyze open and close vowels in terms of their production. It is shown that a dynamic model of lip and mandible interaction can be constructed that predicts the observed durational difference. The assumptions on which these predictions have to be based are then exposed to data on lip and mandible movement.

Open-close dimension of vowels

Traditionally the degree of opening of a vowel is said to be determined by its "tongue height". It has been recognized, however, that the depression of the mandible can also serve as a criterion of openness. But since it is easy to demonstrate that perceptually acceptable
vowels can be produced also with a pencil between the teeth, tongue height rather than jaw opening has been considered the primary feature controlled by the talker.

Nevertheless, a fairly good agreement between classifications based on tongue height and jaw opening is demonstrated by the results of the following experiment: Three subjects were asked to sustain Swedish vowels (long and short). A photographic method was used to record the position of the mandible which was indicated by a special device (3).

Fig. 1-A-1 shows the depression of the mandible as a function of the three natural classes of front, unrounded and front, rounded and back, rounded vowels. Each point represents the mean of approximately 20 measurements from the three subjects. Lines have been drawn to connect vowels assigned to three classes of opening. These plots bear a rather close resemblance to classical vowel quadrilaterals.

Vowel duration and mandible position

Since the inherent duration of a vowel has been associated with its tongue height, and tongue height and jaw opening appear to be correlated in non-compensatory modes of pronunciation, it is natural to expect at least an approximate correlation also between vowel duration and mandible position. Two talkers who participated in the above-mentioned experiment also read lists consisting of randomized sequences of nonsense words containing long and short Swedish vowels in an [I’b-b(b)I] frame. For each talker, the duration of the vowel segment is defined as the interval between the plosion of the first [b] and the initiation of the occlusion for the second [b]. The results of these measurements (pooled and averaged) are presented in Fig. 1-A-2 and compared with data from the first experiment. It is clear that the expected correlation, although gross, is present in these data.

A dynamic model of labial articulation

An articulatory interpretation of the vowel duration measurements of Fig. 1-A-2 where the vowels occurred in the context of [I’b-b(b)I] can be made most easily in terms of the midsagittal separation of
Fig. 1-A. 1. The depression of the mandible as a function of the three natural classes of front, unrounded [-grave, -flat] and front, rounded [-grave, +flat] and back, rounded [+grave, +flat] vowels. Each symbol is based on a pooled average of at least 20 measurements from three talkers.
Fig. I-A-2. Relation between vowel duration in the context of 
[ɪˈb-ʊ(ʊ)l] and degree of jaw opening for two talkers. 
The vowel duration data are pooled averages of 20 
measurements for each vowel. The jaw data are 
based on approximately 15 measurements and are 
from the experiment associated with Fig. I-A-1.
the lips: When this parameter is zero we have a bilabial occlusion; when the parameter assumes a positive value we have an open labial configuration as during a vowel. Consider a point located midsagit-tally on the upper lip near say the vermilion border and a corresponding point on the lower lip. In the following we shall observe lip movements in terms of the displacements of these points. In relation to some reference, for example the maxilla, the displacement of the upper lip is due mainly to labiomuscular forces. The lower lip, on the other hand, moves owing to the presence of labial and/or mandibular muscular forces.

Assume that the upper and lower lips and the mandible can each be represented by a damped spring-mass system \(^{(6)}\). For such a system the displacement \(x(t)\) is governed by the standard equation of motion:

\[
m\ddot{x} + b\dot{x} + kx = f(t) \tag{1}
\]

where \(m\), \(b\), and \(k\) are the mass, friction and spring constants and \(f(t)\) the driving force. Dots above \(x\) represent differentiation with respect to the time variable \(t\). Eq. (1) could for instance be applied to the upper lip. The lower lip and the mandible on the other hand are coupled and thus the differential equations describing their motion differ from Eq. (1). The interaction between the two bodies becomes negligible under certain physical conditions, however. Thus, if the mass of the mandibular structures is assumed to be large, and its resonance frequency small in comparison with the corresponding attributes of the lower lip, the motion of the mandible will in a first approximation be unaffected by the lip and the movement of the lower lip will be the sum of the individual lip and jaw components. Since it appears justified to suppose that in a first approximation, the lips and the mandible come reasonably close to meeting these assumptions under physiological conditions we define \(s(t)\), the midsagittal separation of the lips, as

\[
s(t) = x_u(t) - x_L(t) - x_j(t) - x_0 \tag{2}
\]

This equation is by definition valid only when \(x_u(t) - x_L(t) - x_j(t) - x_0 \geq 0\). For \(x_u(t) - x_L(t) - x_j(t) - x_0 < 0\) we set \(s(t) = 0\). The terms \(x_u\) and \(x_L\) represent the labio-muscular displacements of the upper and lower lips and \(x_j\) denotes the jaw-dependent
depression of the lower lip relative to the neutral location of the upper
lip and $t$ is time. In accordance with the preceding discussion $x_u$ and
$x_j$ are here treated as independent additive terms. These parameters
are illustrated in Fig. I-A-3. Elevation above the neutral or jaw-
dependent position makes a parameter positive and depression makes
it negative. Before insertion into Eq. (2) the signs of $x_u$, $x_\ell$ and $x_j$
in Fig. I-A-3 are thus negative, positive and negative, respectively.
The term $x_o$ is a constant which stands for the uncompressed length
of the tissue between the points on the lips at the moment when $s$ be-
comes zero or the lips just touch. Evidently $x_u - x_\ell - x_j > x_o$ when
the mouth is open, and for closure: $x_u - x_\ell - x_j \leq x_o$. These situ-
ations are depicted in the two halves of Fig. I-A-4. The fact that we al-
low $x_u - x_\ell - x_j$ or the distance between the points on the lips to be-
come smaller than $x_o$ presupposes a certain compressibility of their
marginal tissue. In the right half of Fig. I-A-4 this circumstance is
taken into account by the presence of the spring connecting the two
bodies. In the argument to follow below this element will be assumed
to be zero but have the length of $x_o$.

Simulation of vowel duration

The model developed above and illustrated in Figs. I-A-3 and
I-A-4 can now be used to compute the hypothetical course of the mid-
sagittal separation of the "lips", $s(t)$, for say, [bVb] syllables to
which we assign various degrees of vowel or jaw opening. In these
computations we shall rely on a repeated application of Eq. (1) which
will enable us to derive the component motions of $s(t)$, viz., $x_u(t)$,
$x_\ell(t)$ and $x_j(t)$. From Eq. (2) we shall then obtain the time functions
representing $s(t)$ which can be studied with respect to "acoustic vowel"
duration as a function of "jaw" opening.

Our first step is to assign certain "labial" and "mandibular" target
positions to each segment of [bVb]. For [b] the upper "lip" is de-
pressed to make contact with the lower "lip". For the vowel it is
raised. The lower "lip" is elevated for [b] and lowered for the vowel.
The mandible is raised for [b] and assumes three degrees of opening
for the V which is to be simulated as a "close", "half-open", and
"open" vowel. The notion of articulatory target position is implicit
in traditional phonetics which associates with each phoneme in a given
Fig. I-A-3. Graphical illustration of labial and mandibular parameters of \( s \), the midsagittal separation of the lips. \( x_u \) and \( x_j \) represent the labio-muscular displacement of the upper and lower lips relative to the maxilla and the mandible respectively, \( x_j \) denotes the jaw-dependent displacement of the lower lip relative to the neutral location of the upper lip. The combined distances of \( x_u \) and \( x_j \) from the lip margins are represented by \( s \). The arrows indicate the direction of the displacement. Upward displacements occur in the positive \( x \)-direction. For the particular labial configuration schematized in this figure the values of \( x_u \) and \( x_j \) are thus negative whereas \( x_j \) is positive.
Fig. I-A-4. The parameters and configuration of Fig. I-A-3 interpreted in terms of mechanical models of the lips. Upward displacements occur in the positive x-direction. $x_u$ is the displacement from a neutral position that results from applying a force to the upper "lip". $x_j$ is the "jaw"-dependent position of the lower "lip" and $x_j$ is the displacement away from this position that is due to the application of a "labial" force. In the left part the systems are independent; in the right half they interact. The constant $x_u$ is equal to the uncompressed length of the spring connecting the two bodies in the right part of the figure.

\[ I. x_u - x_j > x_0 \]
\[ II. x_u - x_j \leq x_0 \]
language certain ideal positions of the articulatory organs and interprets syllables, words, phrases etc. basically as sequences of movements between such positions (cf. Sievers's "Stellungslauten und Übergangslauten" (8)). For a given articulatory component to reach its target positions in a sequence such as [bVb] a time-varying force must be applied to it. For simplicity we make the value of this force constant for the duration of each phoneme. As a result the syllabic pattern of the force becomes rectangular in shape. The rectangular pulses shown in the left part of Fig. I-A-5 represent such stylized forces applied to the upper and lower "lips" and the "mandible" in [bVb]. In terms of these patterns "vowel duration" can be defined "articulatorily" as the interval during which the force assumes the value appropriate for the 'vowel.' In Fig. I-A-5 this interval is denoted D, for the "lips" and D, for the "mandible". The response of the "lips", or the "mandible", to any given input force, f(t), can be determined from Eq. (1) for instance by means of Laplace transforms (7, 9) once the values of the constants m, b, and k have been chosen. For computational convenience b was set equal to 2√km in deriving the displacements of these systems. This choice makes them critically damped and gives them an impulse response of te^-αt where α = √k/m. The general form of the response of a critically damped and uncoupled second-order system to a rectangular force pulse is

\[ x(t) = A_o + Au(t-a)[1 - α(t-a)e^{-α(t-a)} - e^{-α(t-a)}] - \\
- Au(t-b)[1 - α(t-b)e^{-α(t-b)} - e^{-α(t-b)}] \]  

(3)

where \( A_o \) is the initial location of the system; \( A \) the total extent of displacement; \( α \) determines the rate of movement; \( t \) is time and \( u(t-a) \) and \( u(t-b) \) represent the conventional notation of unit step that occur at \( t=a \) and \( t=b \) respectively. By definition these functions assume the values of zero for \( (t-a) < 0 \) and \( (t-b) < 0 \) and one for \( (t-a) \geq 0 \) and \( (t-b) \geq 0 \). In the present application \( t=a \) and \( t=b \) stand for the moments of onset of "phoneme commands". Eq. (3) has been tested empirically and found quite realistic (10). (As an alternative to Eq. (3) the response of an overdamped second-order system might also be considered.) The curves in the right half of Fig. I-A-4 all have the form specified by Eq. (3) and represent responses to the
Fig. 1-A-5. Time-variations of input forces to the model of Fig. 1-A-4 for a hypothetical [bVb] syllable (left part). The model components are taken to be second-order and non-oscillatory systems. The upper lip system responds to the input force as shown in the right half of the figure. The displacement of the lower lip contains the labio-muscular and a jaw-dependent component indicated below to the right. Three degrees of vowel opening are reflected in the mandibular and jaw-dependent curves.
force patterns shown on the left of this figure. The functions describing the jaw-dependent depression of the lower "lip", \(x_j(t)\), have the same form as the underlying displacements of the "mandible" itself which are not shown. The value of \(\alpha\) is larger for the "lips" than for the "jaw". That is, the "jaw" is inherently slower than the "lips".

To derive \(s(t)\) it is also necessary to make an assumption about the relative timing of the labio-muscular components \(x_u(t)\) and \(x_j(t)\) on the one hand, and the jaw-dependent component \(x_j(t)\) on the other. We shall assume that the coordination of the lips and the mandible follows the same coarticulatory pattern as the tongue and the lips generally do in a syllable such as for instance [tu] in which the labialization for the vowel is known to develop during the [t]-closure in many languages. In accordance with this pattern the mandibular opening for the vowel in [bVb] should be anticipated during the occlusion of the initial [b]. This assumption is supported by data on labial and mandibular articulation reported by Fujimura (11) and indirectly by the work of Öhman on coarticulation (12, 13). Fujimura notes in discussing his measurements on bilabial occlusive consonants that "the jaw is already lowered considerably when the plosion occurs. The movement of the jaw after the plosion is considerably slower than that of the lips". As can be seen in the left part of Fig. I-A-5 the "mandibular" force pulse leads those of the "lips".

The result of combining according to Eq. (2) the time functions derived so far and determining \(x_0\) is shown in Fig. I-A-6. The solid segments of the curves represent the "acoustic vowel" segments or the intervals during which \(x_u - x_j > x_0\). It can be seen that as the vowel opening changes from "close" to "half-open" and to "open" the distance between the "zero-crossings" of \(s(t)\) increases and acoustic vowel duration, \(D_Y\) consequently also increases. This variability has two causes evidently. There is coarticulation during the initial [b] for the "mandible" starts maneuvering towards the vowel already during the "labial" closure. There is also coarticulation during the final [b] in that the "mandible" approaches its target position for the [b]-closure so slowly that the moment of "labial" contact must occur as a function of the extent of movement from the vowel target.
To summarize, a dynamic model of labial and mandibular coordination can be constructed in such a way that it correctly predicts a durational difference between open and close vowels in \([b\tilde{v}b]\) type of syllables. This result is based on a number of simplifying assumptions concerning the shape of the input forces to the lip and jaw systems and the dynamic properties of these systems. It is clear that further constraints must be imposed on the choice of time constants, impulse response etc. before the empirically observed correlation between vowel duration and mandible position can be predicted with quantitative precision. The major purpose of Fig. I-A-5 and Fig. I-A-6 and the associated discussion, however, has been to demonstrate that the durational variance of open and close vowels is not necessarily a characteristic of the input control signals to the model. Neither \(D_j\) nor \(D_k\) were varied as a function of vowel opening. It can be generated simply as a result of the superposition of mandibular movement on the opening and closing gestures of the lips. These findings are in qualitative agreement with the data of Fig. I-A-2 and suggest the hypothesis that the variability of the duration of open and close vowels should be attributed to mandibular coarticulation. It now remains to be seen to what extent this hypothesis is compatible also with data on lip and mandible movement.

Experimental method and procedures

A procedure has been developed to record lip and jaw movements continuously and in synchrony with the speech signal. This procedure which is an improved version of the method described earlier by Lindblom and Bivner (14) involves the use of cylindrical miniature lamps which are about 0.5 mm in diameter and 3.0 mm long, and are attached to the talker's lips and to a special device indicating the position of the mandible. Similar techniques have been described earlier by Jeaffreson (15) and more recently by Kozhevnikov and Chistovich (16). The movement of the lamps during speech is registered by a 35 mm oscilloscope camera in which the film runs perpendicular to the movement dimension under analysis.

The camera was run at a speed of approximately 2 cm/sec. Since the noise of the camera was considerable it was placed behind a sound-insulating door in a room adjacent to that used by the subject
Fig. I-A-6. The course of $s(t)$, the midsagittal separation of the lips in [bVb] syllable simulated for three degrees of vowel opening. Acoustic vowel duration, $D_v$, can be defined as the interval during which $s > 0$. This interval is seen to increase as a function of vowel opening.
and the experimenters. The light from the lamps reached the camera by way of a circular window in the door. Thanks to this arrangement a signal-to-noise ratio better than 40 dB was measured on the tape recordings. Spectrograms of speech recorded during the experimental sessions showed no trace of interfering noise, Fig. I-A-7.

To minimize the superposition of head and body movements on the film records, a head clamp device was attached to the sound-insulating door. This device consisted of a curved metal bar. In the present experiments it was oriented so that the subject faced the camera when he pressed his forehead against it. Its distance to the camera could be adjusted along two firm horizontal metal rods. These rods were parallel to the main axis of the camera lens system and could thus serve as references when a subject was positioned in relation to the camera. Anatomical reference lines such as the Frankfurt horizontal (17, 18) were marked on the subject's head to facilitate a suitable and reproducible relative positioning. To fix the subject's head against the head rest, leather straps with an adjustable buckle were also used.

In the experiments to be reported below measurements were made of the vertical movements of the lips and the jaw. Lamps were attached midsagittally to the lips and a device for indicating the mandible. Since slight head and body movements could not be entirely eliminated in spite of the above-mentioned measures it proved desirable to introduce a reference lamp that would follow such movements. This lamp was located just above the upper lip in approximately the same coronal plane as the lips and the mandible lamps. This was arranged with the aid of a pair of tight-fitting glasses provided with a trunk-like extension for holding the lamp. The lip lamps were attached by means of a drop of glue only insignificantly larger than the contact area and in such a way that they would always be separated even during a [b]-closure. The cables leading to the lamps were thin and extra flexible. It is clear that the skin of the chin can move considerably in relation to the mandible as the muscles of the lips contract. Consequently it cannot be used as an accurate indication of mandible position. For this purpose dental casts were made for each subject individually and from these, cap splints. The cap splints were thin and light and fitted tightly to the lower incisors.
Fig. I-A-7. Spectrogram of [I'ba:bl]. At top the synchronization pulses.
A piece of firm wire was attached midsagittally and ventrally to each such device. It was adjusted so that the lips would touch it only in bilabial stops for which its vertical location would coincide roughly with the level of lip contact.

Before the sessions the subjects had plenty of time to get used to talking with the lamps and the jaw device on. After they had had some short practice it was not possible to hear whether they were talking with or without lamps and cap splint. Listening to tape-recordings of the test lists confirms that impression. Moreover, film records of words from these lists pronounced with and without the cap splint fail to reveal radical differences between the two conditions. (Fig. I-A-8). The vertical distance between the lamps was measured for three talkers and for both conditions in sustained versions of [i:] and [a:] which occur in the present speech materials. The differences fail to reach significance at the 5% level. In Fig. I-A-8 a comparison is made between normalized and averaged records of the vertical distance between the lip lamps to assess the effect of the cap splint on the movement of the lamps in [I'bi:bI]. As can be seen the curves for the two conditions are almost identical. The normalization and averaging procedure will be described below in the section on Experimental Results.

For synchronization purposes the current through the lamps was pulsed. The speech signal would thus be recorded on one of the channels of a twin-track tape-recorder the other channel being used for the pulses. The lamps were lit by the experimenter for each item in the test list individually. To facilitate the detailed temporal alignment of the acoustic and articulatory records every eighth pulse differed from the rest in shape and would produce a thicker trace on the film. In Fig. I-A-7 the synchronization pulses appear at the top of the spectrogram. The shape of the pulses was chosen so as to speed up the onset characteristics of the lamps. The delay between the onset of a given pulse and the onset of the corresponding photographic trace was thus rendered negligible. The decay characteristics of the lamps permitted no faster rate than 50 sec\(^{-1}\) for the pulses to appear distinctly on the film. In view of the relatively slow movements of the lips and the jaw this was considered quite sufficient.
DISTANCE BETWEEN LIP LAMPS (mm)

Fig. 1. A-B. Effect of jaw device on separation of lip lamps.
It was clear that the movements of the lips and the mandible might not always take place at right angles with the film path. Often such movements are also transversal. A lamp can accordingly be displaced horizontally away from its ideal vertical path, that is, along the time scale of the film and consequently distort the time relations among the traces. One valuable feature of using pulsed instead of continuous traces is that such distortions can be corrected for. Fig. I-A-9 shows an example of the type of record that can be obtained with the present procedures.

During a typical recording session the subject would be seated in a darkened quiet room reading the test words from spotlit sheets of paper. Three subjects, all laboratory employees, were used. Two of these (subjects A and C) had had no training before as experimental subjects and were not aware of the purpose of the experiments. The third talker (subject B) was the present writer. All three have normal speech and speak a Stockholm variety of Swedish. They all have normal occlusion (Angle, class I). Talker A has a slightly open bite. They also served as subjects in the experiment of Fig. I-A-1.

The speech materials recorded contain the four nonsense words [I'bax:bI], [I'bi:bI], [I'babbI], and [I'blbbl] which were pronounced in the context of [sej ___ I'jë:n] (say ___ again). The absolute level of the speech signal one meter from the talkers' lips was of the order of 70 dB rel. 0.0002 dyn/cm². After about a dozen items had been read from the list the recording was interrupted and the talker would be "unchained" for a short while. Before each new recording care was taken to reproduce the standard positioning so that the Frankfurt horizontal would be parallel with the axis of the camera lens system. This procedure was repeated until a sufficient number of items had been recorded. On the films a given sequence of test words was preceded by a recording of the mandible in a maximally elevated position so that its location in relation to the maxilla could be inferred. The degree of jaw opening as used in Figs. I-A-1 and I-A-2 thus refers to the vertical projection of the distance of the jaw lamp from its position for closed jaws. To isolate the jaw-dependence of the distance between the lamps from its
Fig. I-A-9. Example of film record of articulatory movements. The traces marked A, B, C and D pertain to the reference lamp, the upper lip, lower lip and jaw lamps respectively. The utterance is [I'bi:bI].
labio-muscular dependence the subjects also opened and closed their mouths slowly with as relaxed lips as possible. Individual plots of the separation of the lamps $[x_u - x_j - x_j']$ where $x_u = x_j = 0$ as a function of jaw position indicate linear relationships with proportionality constants not markedly different from one. The total scatter around lines fitted by visual inspection to these plots was found to be approximately ± 2.5 mm.

From past experience it is clear that the described method compares well with alternative techniques such as high-speed\textsuperscript{(19)} and cineradiographic\textsuperscript{(20)} photography as regards accuracy of measurement. A definite advantage is that the time variations of individual parameters can be obtained without a time-consuming examination of individual film frames. On the other hand, it shares many disadvantages with other articulatory and physiological measurements which all require long-winded preparations and tenacity on the part of the subject.

**Experimental results**

**Quality of data**

The technique described above was used to acquire lip and jaw data on the production of open and close vowels. These data were obtained from enlarged tracings of records such as Fig. I-A-9. The accuracy with which spatial coordinates can be measured on the film is limited chiefly by the enlargement factor and the resolution of the light pulses. During fast movements the pulses would sometimes appear weaker and slightly blurred. This difficulty occurred only occasionally and the error of a given measurement is estimated to be within a fraction of a millimeter in most cases. Since the pulse rate was known measurements of time coordinates presented no problem. In synchronizing the traces care was taken to correct for distortions and spurious translations of the time scales caused by transversal movements, faulty midsagittal alignment of lamps etc. The temporal alignment of acoustic and articulatory records should be accurate within a few milliseconds. The identification of acoustic segment boundaries was made from mingograms since it was found that the segmentation of stop-vowel stop sequences on such records gave as accurate result as that based on wide-band spectrograms.
Normalization and presentation of data

Ten repetitions of each test word were analyzed for each talker. Since the timing of movements varied somewhat from one utterance to the next averaging could not be undertaken until the time scales of each utterance had been normalized. The normalization procedure adopted was as follows. Each curve representing the distance between the lip lamps was divided into four time segments the location of a segment boundary being determined by the points of minimum rate of change. The time scale of each segment would then be multiplied by a normalization factor in such a way that the duration of the segment in question would coincide with the average duration of that segment. This procedure produced standard deviations that were generally within a millimeter at places where the rate of change is slow but slightly larger at other points on the curve.

The four test words [I bæ : bI], [I babbI], [I bɪ bɪ bI], and [I bɪ bɪ bɪ bI] were evaluated for each of the three talkers with respect to
(a) the separation of the lips or, in terms of the model parameters: $x_u - x_j^0 - x_j^0$;
(b) the jaw-dependent depression of the lower lip, $x_j^0$;
(c) the labio-muscular components of lip separation, $x_u - x_u^0$.

In Figs. I-A-10 - I-A-15, a, b, and c are shown in the top, middle and bottom parts. The separation of the lips, $x_u - x_u^0 - x_j^0$, is the difference between curve B and curve C in Fig. I-A-9. The position of the mandible is indicated by the distance between curves A and D. Its position relative to the maxilla, or the degree of jaw opening, is defined as the distance between the A and D curves minus the difference between these curves for closed jaws. The separation of the lips as defined above reduces to $-x_j^0$ when $x_u = x_u^0 = 0$, that is, when the lip muscles are relaxed. Data on $x_j^0$ as a function of mandible position were obtained by recording each subject slowly moving only his jaw while keeping his lips relaxed. Plots were prepared for each subject individually to permit the determination of $x_j^0$ for any given position of the mandible. Since these plots indicated approximately linear relationships whose slope approached 1 for all three talkers the curves labeled $x_j^0$ can be interpreted to describe also the shape of the jaw movement in a first approximation. The "labio-muscular" components of lip separation $x_u - x_u^0$ are obtained by adding the top and middle curves of Figs. I-A-10 - I-A-15.
In these figures comparisons are presented between [I'ba:bI] and [I'bi:bI] and between [I'babbl] and [I'blbbl]. The vertical lines represent plosions and closures for [b]. For words with a close stressed vowel dashed lines are used to indicate these boundaries. For the words containing [a:] or [a] solid lines are used. From left to right the lines stand for the moments of closure for the first [b], release for the first [b], closure for the second [b], and release for the second [b]. The time reference common to each pair of utterances is the beginning of the occlusion for the initial [b].

The time location of a given boundary corresponds to the moment at which \( x_u - x_b - x_j \) is equal to its average value for this boundary.

**Talker A: [a:] - [i:], Fig. I-A-10**

The top part of Fig. I-A-10 shows the distance between the lip lamps, or \( x_u - x_b - x_j \), in [I'ba:bI] and [I'bi:bI]. The curves vary smoothly through the entire words exhibiting maxima of separation during the vowels and minima during the consonants. The minimum for the second [b] is somewhat deeper for [I'bi:bI]. For this word \( x_u - x_b - x_j \) assumes a larger value at the first release than at the closure for the second [b]. It is also larger for [a:] than for [i:] at the latter moment. A general feature of the curves marked \( x_u - x_b - x_j \) is that, during the [b] segments, these curves continue to decrease till a minimum is reached at a point close to the middle of the segment. Evidently the compressibility of the lip margins prevents \( x_u - x_b - x_j \) from becoming "clipped" during the closure intervals.

An examination of \( x_j \) indicates that the mandibular position for [a:] is anticipated already during the occlusion of the first [b]. For [i:] \( x_j \) shows but little opening excursion. Returning to its position for the second [b] \( x_j \) changes slowly. As a result the curves for [I'ba:bI] and [I'bi:bI] differ considerably at the beginning of occlusion of the final [b].

The bottom part of Fig. I-A-10 shows the curves corresponding to \( x_u - x_b \) which are similar with respect to the negative minimum values reached during the [b] segments and the positive maxima attained during the vowels.
Talker A: [a] - [I], Fig. I-A-11

The curves labeled $x_u - x_j$ differ mainly as regards opening for the stressed vowel. Also the [a] is longer than [I] whereas the consonant segments have about the same duration. The separation of the lips is larger at the instant of release than at moment of occlusion (Table I-A-Ia and b).

In terms of $x_j$ the opening gesture for [a] seems to have been initiated well ahead of the plosion. This appears to be true also for the [I] curve in which the point of maximum excursion is reached somewhat earlier than that for [a]. The movement from the vowel [a] towards the final [b] progresses slowly. As a result the [b] following [a] is delayed.

For $x_u - x_j$ marked differences can be observed between [I'babbI] and [I'bbbbl]. The [I] curve begins to approach its positive value for the stressed vowel earlier than the [a] curve which also displays a somewhat deeper minimum for the second [b]. Testing for significance in the intervals where the curves differ indicates highly significant differences in the regions of both the first and the second [b] occlusions. These tests were performed not on curve segments but only for pairs of ordinates at individual time coordinates. Since fewer degrees of freedom are involved in tests of the latter kind we are justified in concluding that not only the tested pairs but also the curves are likely to differ significantly in these intervals.

Talker B: [a:] - [i:], Fig. I-A-12

Regarding $x_u - x_j$ we note that it is larger at the moments of release than at the beginning of closure.

For [a:] $x_j$ starts approaching the stressed vowel position during the first [b] and reaches its value for the second [b] considerably later than the initiation of the occlusion.

Talker B: [a] - [I], Fig. I-A-13

$x_u - x_j$ is larger at the plosions than at the moments of closure.
Fig. I-A-11. Comparison of labial movement components for [I' babbI] and [I' bibbI]. Talker A.
Fig. I-A-12. Comparison of labial movement components for [I’bɑːbiː] and [I’bitbiː]. Talker B.
Fig. I-A-13. Comparison of labial movement components for [I'babbi] and [I'bUbbi]. Talker B.
It can be inferred from \( x_j \) that the jaw movement for [a] begins during the first occlusion and that the return movement is completed not until during the final [b] occlusion.

In neither of the figures does talker B show the reorganization of \( x_u - x_j \) observed for talker A in Fig. I-A-11. Furthermore, the differences between the open and close vowels as reflected in \( x_j \) are somewhat smaller than for talker A, especially at the moments of closure for the final [b]'s.

**Talker C: [a:] - [i:], Fig. I-A-14**

The duration of the consonant segments is comparable in the two words. [a:] is longer than [i:]. There is again the tendency for the lip lamps to be more widely separated at the release than at the beginning of closure.

The \( x_j \) curve for [I'babbl] exhibits considerable excursion for the stressed vowel. There is anticipation of the [a:] position during the initial occlusion and the attainment of the position for the second [b] occurs at a slow rate as observed also in the previously examined utterances.

Considering \( x_u - x_j \) we see that this component reaches a value of about 7-8 mm during [i:] but about 2 mm for [a: ]. In Swedish [a:] is a slightly rounded vowel whereas [i:] is spread. Significance tests performed as described above indicate a high probability that the dashed and solid curves are different in the neighborhood of the [b] releases. The relative timing of curves for [a:] and for [i:] and for the final vowel [I] is similar to that observed earlier for talker A in [I'babbl] and [I'bIbbI]. The solid curve seems to be delayed from the first [b] occlusion onward.

**Talker C: [a] - [I], Fig. I-A-15**

The short vowels for this talker produce data approximately analogous to those for [a:] - [i:]. Thus the consonant segments are not markedly different in duration. [a] is longer than [I]. Again, \( x_u - x_j \) tends to be larger at the plosions than at the beginning of closure.
Fig. I-A-14. Comparison of labial movement components
[I'ba:bl] and [I'bi:bl]. Talker C.
Fig. I-A-15. Comparison of labial movement components for [I'babbi] and [I'blbbi]. Talker C.
In the motion towards [a] is initiated already during the first occlusion and the value for the final [b] is attained only during the later half of the occlusive segment.

The functions representing \( x_u - x_\lambda \) differ significantly in the regions of the [b] releases. Also in these curves some of the events for [I] occur earlier than for [a].

**Summary of results**

1. The coordination of labial and mandibular gestures is characterized by coarticulation. The mandible begins its opening movement for the stressed vowel in [I' bV(:)b(b)I] while the lips are still in position for the [b]. Similarly it completes its closing movement for the second [b] after the labial closure for this consonant has been attained. These effects are present for all the talkers.

2. Comparing the words with open and close vowel pairs it is seen that there is reorganization of the "labio-muscular" components of lip separation \( x_u - x_\lambda \). This effect is clearly present only in Figs, I-A-11, I-A-14, I-A-15.

3. The distance between the lip lamps is not the same for plosions and closure onsets. It tends to be larger at the moment of releases. It is also larger for [a:] at the closure of [b]. (Table I-A-Ia and I-A-Ib)

4. Open vowels are longer than close vowels in all cases. However, the differences are in some cases somewhat smaller than indicated by the data in Fig. I-A-2.

**TABLE I-A-Ia.** Distance in mm between lamps at the moment of closure of [b].

<table>
<thead>
<tr>
<th>Preceding vowel</th>
<th>Talker:</th>
<th>row mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>['a:]</td>
<td>12.5</td>
<td>12</td>
</tr>
<tr>
<td>['i:]</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>['a]</td>
<td>8.5</td>
<td>9.5</td>
</tr>
<tr>
<td>['I]</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>[I]</td>
<td>8</td>
<td>10.5</td>
</tr>
<tr>
<td><strong>column mean</strong></td>
<td>8.5</td>
<td>10.5</td>
</tr>
</tbody>
</table>
TABLE I-A-Ib. Distance in mm between lamps at the moment of plosion of [b].

<table>
<thead>
<tr>
<th>Following vowel</th>
<th>Talker:</th>
<th>row mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>[’a:]</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>[’i:]</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>[’a]</td>
<td>14</td>
<td>14.5</td>
</tr>
<tr>
<td>[’I]</td>
<td>13.5</td>
<td>13.5</td>
</tr>
<tr>
<td>[I]</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

column mean: 13, 13.5, 15

Interpretation of \( x_0 \). There are two tendencies in Table I-A-I that require an explanation. Firstly, the separation of the lip lamps is larger at plosions than at the beginning of closures. This discrepancy should probably be attributed to the presence of adhesive forces that make the lips stick together and appear as soon as the lips make contact or as the opening of the lips is begun. On the termination of these forces discontinuities can sometimes be seen in the individual lip traces at the moments of release. In these cases it is possible that oscillations also occur as found in previous investigations \(^{14,21}\).

Our earlier experience of high-speed film analysis of the lips \(^{22}\) convinces us that such adhesion effects are by no means uncommon under normal speaking conditions and they should not be attributed exclusively to the present use of lamps and the special jaw device. The accommodation of these facts by the model of Fig. I-A-4 is achieved most simply by supplementing the spring connection the "lip" masses by an element of viscous damping which is introduced as the driving forces applied to the lips change signs and become opening forces. Two values must be assigned to the constant \( x_0 \) of Eq. \((2)\): one for closure and one for release. In this revised form the present lip model bears a certain resemblance to Flanagan's model of vocal fold vibration \(^{23}\) in which a friction component is discretely introduced during vocal fold contact.

Secondly, at the beginning of occlusion \( x_u - x_L - x_j \) tends to be larger for [’a:] than for the other vowels. This trend recurs for all three talkers. As remarked earlier Swedish [’a:] is a somewhat
rounded vowel whereas [ɪ], [a], and [ɪ] are spread. If the [b] closure after [ɑː] were made with slightly rounded lips and the closure after the other vowels with spread lips it seems not unlikely that there should be some protrusion following [ɑː] but spreading and stretching following the others. When protruded the lip margins would thus tend to thicken and this would have the effect of permitting the closure of the lips with somewhat more widely separated lip lamps. When stretched and spread the lip margins would tend to become narrower and thinner and closure would be made with the lips at a smaller distance from each other. Possibly these coarticulation effects offer one way of accounting for the vowel dependent variations of $x_u - x_\ell - x_j$ in Table I-A-I.

**Interpretation of $x_j$.** In the development of the model we defined $x_j$ as the jaw-dependent depression of the lower lip measured in relation to the neutral location of the upper lip (Fig. I-A-3). This parameter consequently specifies the position of the lower lip that would have been observed had there been no labio-muscular displacement but only mandibular movement. The shape of $x_j(t)$ is more or less that of the jaw movement. The behavior of this function in Figs. I-A-10 - I-A-15 confirms the assumptions concerning coarticulation that had to be made in using the lip-mandible model to predict vowel duration. An examination of $x_j$ at the moment of release in words with close stressed vowels reveals that the contribution of the jaw to the position of the lower lip is on the average about 3 mm larger for open vowels than for the close cognates. These differences should, but does not always, favor an earlier release of [b] for open than for close vowels. At closure onset for [b] preceded by [ɪ] or [ɪ] the contribution of the jaw movement is on the average 8-9 mm larger for the corresponding open vowel (Talkers A and C). For talker B this value is about 4 mm. This circumstance should have the effect of delaying the instant of closure after an open vowel. So it does in all cases as seen from the relative timings of the final [b] closure.

**Interpretation of $x_u - x_\ell$**

$x_u$ represents the displacement of the upper lip relative to the maxilla. $x_\ell$ can be looked upon as the displacement of the lower lip relative to the mandible. The extraction of the jaw-dependent
component \( x_j \) from \( x_u - x_\ell - x_j \) might leave a residue, \( x_u - x_\ell' \), whose shape would partly be influenced by adhesion effects and other mechanical factors. In a first approximation it seems reasonable to interpret these parameters as primarily labio-muscular, that is, they reflect the shortening of the lip muscles. When \( x_u - x_\ell > 0 \) the total separation of the lips will exceed their neutral separation determined by the position of the mandible. This situation may occur because the upper lip has been raised, lower lip lowered or both. When \( x_u - x_\ell = 0 \) the separation of the lips is equal to its jaw-dependent value. When \( x_u - x_\ell < 0 \) the action of the labial muscles make the separation between the lips smaller than when it is neutral. This might be caused by the lowering of the upper lip, the elevation of the lower lip or both.

In Figs. I-A-10 - I-A-15 the general course of \( x_u - x_\ell \) is similar for all utterances. Positive maxima are reached during vowels and negative maxima during the consonants. The interpretation of these curves is straightforward. During the vowels the lip separation is wider than it is when neutral. The upper lip may accordingly be raised and the lower lip depressed. Rounding and spreading, although different mechanisms, produce the same result in this parameter as seen from a comparison between the values during \([a:]\) and \([i:]\) for talkers A and B. During the consonants \( x_u - x_\ell \) goes negative. Inspection of the individual traces for the upper and lower lips indicates that the lips approach each other to make the contact for \([b]\) the upper lip being depressed and the lower lip elevated. These displacements result in a distance between the lips which is smaller than the jaw-dependent separation.

In Figs. I-A-11, I-A-14, and I-A-15 which pertain to talkers A and C there are large differences between the timing of the \( x_u - x_\ell \) gestures from the initial \([b]'\)s into the stressed vowels. The phase relation is such that the close vowel curve leads that for the open vowel. The timing difference is large enough to prevent the acoustic \([b]\) segments from becoming shorter in duration before the open vowels and thus also to prevent these vowels from becoming still longer. This point can be further illustrated as follows. Suppose that for \([I'b\alpha:bI]\) of Fig. I-A-14 we compute a hybrid version of \( x_u - x_\ell - x_j \) whose \( x_j \) is that of \([I'ba:bI]\) but whose \( x_u - x_\ell \) is that of \([I'bi:bI]\). What effect would this have on the
time location of the boundary between the initial [b] and [a:]? If the $x_o$ value for [i:] in Table I-A-Ib is used the duration of [b] decreases by approximately 45 msec from 120 to 75 msec. The corresponding operations for [l'babbl] (Talkers A and C) gives a reduction of the duration of [b] by 20 and 25 msec respectively. It seems rather unlikely that the space-time patterns of muscular shortening are basically the same in these cases and that some mechanical effect such as adhesion could account for the differences in $x_u - x_j$. Rather, it looks as if the opening of the initial [b] is made with the lips when [i:] or [I] follow but to a larger extent with the jaw when [a:] or [a] follow. In the region of the initial [b] the course of $x_u - x_j - x_j$ remains the same for each open close pair. The component gestures seem to be shaped with respect to the total movement. Note also that $x_u - x_j$ exhibits significantly deeper minima for the final [b] following the open vowels (Figs. I-A-11, I-A-14, and I-A-15). The deeper minima imply that the lips strive to approach each other more closely during the closing movement from the open vowels thus compensating for the large separation caused by the jaw. There is evidently adjustment of lip activity but not of the movement of the more massive mandibular structures.

These compensatory modes of articulatory behavior raise the question concerning the involvement of sensory feedback in the motor organization of speech production. It is well known that the control of human and animal movement is extensively based on this principle (24). It has been shown that reflex action of the lips can be elicited by mechanical stimulation (25). This result appears to indicate that the contraction of the orbicularis oris is under afferent control. Furthermore an extensive electromyographic analysis of facial muscle activity during speech has recently been undertaken by Öhman (26) who finds the EMG patterns compatible with an interpretation in terms of general neurophysiological principles of feedback. Moreover, Hosokawa (27) reviewing the literature on the sensory innervation of facial muscles reports one study that professes to have established muscle spindles in the muscles of the lips and in the region of the oral angle. Consequently movements such as those observed in the present study are probably a priori most correctly analyzed in terms of neural circuitry characterized by feedback.
Summary of interpretations: The primary reason why open vowels tend to be longer than close vowels is that the extent of mandibular movement is larger for open vowels even during the occlusion of the initial consonant and that the mandibular off-glide movement from the vowel into the final consonant progresses so slowly that the contact between the articulators for the consonant is delayed in the context of the open vowel. In the present data the timing of the jaw movement is such that the latter effect is the more pronounced.

However, there is some evidence (Figs. I-A-11, I-A-14, and I-A-15) that the lip gestures proper can be reorganized in such a way so as to compensate for the contribution of the jaw movement to the opening of the lips for the initial [b].

In Eq. (2) $x_0$ is one of the determinants of the temporal occurrence of release and closure. Since it was found to assume a larger value after the rounded [a:] than after [i:] this difference contributes also towards reducing the variability of vowel duration that the timing and the sluggishness of the mandible tend to bring about.

In summary, the data demonstrate the realism of the lip-mandible model and the assumptions made initially regarding mandibular timing and rate of movement.

Discussion

Generality of proposed mechanism

The results obtained encourage the belief that the relative lengthening of open vowels that can be observed also in the contexts of [d-d] and [g-g] can be given an explanation similar to that developed above for [b-b]. To make vowel duration dependent on vowel opening for [d] and [g] it will be necessary to construct articulatory models that incorporate separate control of the mandible and the tip and body of the tongue. So far the modeling of articulatory structures has not been attempted in such detail.

There is evidence that the durational variations correlated with vowel opening appear also in vowels followed by [p], [t] or [k]. Before these consonants the acoustic termination of the vowel segment is usually not determined by the onset of the articulatory closure but by the offset of phonatory activity. Sound source features enter as
criteria of the acoustic end of a vowel also before voiceless fricatives. These cases call for a consideration of the temporal relations between phonatory and articulatory gesture initiations as well as of parameters controlling laryngeal vibration which are not contained in the present form of the model, e.g., the pressure drop across the glottis, the position and tension of the vocal folds.

Without further study it is not possible to claim that the coordination mechanism under analysis is a universal mechanism of vowel and syllable production. Coarticulation on which the durational effects are dependent can be looked upon as a temporal overlap of maneuvers towards different but concatenated articulatory goals. By means of this organization the system manages to speed up the actualization of adjacent sounds without having to speed up the rate of movement of the individual component gestures. In view of the allegedly rather slow speed of articulatory movement and of mandibular movement in particular we would like to suggest tentatively, however, that the mechanism studied is likely to be fairly wide-spread among the languages of the world. To a certain extent this expectation is reinforced by the number of languages for which a durational difference between open and close vowels has so far been reported, e.g., American English (28, 29), British English (30), Danish (31), German (32), Hungarian (33), Icelandic (34), Italian (35), Russian (36), Spanish (37), Lappish (38), and Swedish (39). It is true, however, that there are one or two studies that have failed to reveal a difference. For Norwegian, for instance, vowel duration in the environment of voiced and voiceless fricatives was reported not to vary with vowel opening (40). No data for stop contexts were given, however.

Previous explanations

Previous attempts to explain the durational variability of open and close vowels have produced two major hypotheses: the "energy expenditure" hypothesis, and the "articulatory distance" hypothesis. According to the "energy expenditure" hypothesis the temporal organization of speech sounds is determined by the amount of physiological energy that is consumed in producing them. During a close vowel more energy is expended than during an opener one. If the energy per vowel were kept constant [i] would be shorter than [a]. This explanation was first introduced by Meyer (30) and has appeared in related form also in later discussions.
Jespersen (41) suggests that the duration of a vowel is a function of its articulatory "distance" to the adjacent consonants. For this interpretation to work it is obvious that, moreover, the rate of movement must not change appreciably so as to compensate for the distance involved. This hypothesis has been accepted by many investigators. The coordination mechanism studied in this investigation permits the formulation of a special version of it. It is of interest to note that Åima (38) although accepting Meyer's hypothesis suggests also that the jaw movement might play a role in lengthening the open vowels.

Motor control of vowel duration

The model presented above suggests a framework in terms of which labial and mandibular movement can be represented quantitatively. Each component parameter can be quantified with respect to rate, extent of movement and moment of gesture initiation. When it was used to simulate open and close vowels an assumption was made concerning the duration of the vowel commands, $D_2$ and $D_1$, which remained invariant and independent of vowel identity (Fig. I-A-5). In spite of the imposed constraint, however, the model generated variations in acoustic vowel duration similar to those generally observed for many languages (Fig. I-A-6). An implication of this result is that the duration of human open and close vowels might exhibit less variability when considered in terms of the complex of physiological commands that give rise to these sounds. This hypothesis has been propounded also by Fischer-Jørgensen (42) who develops the "articulatory distance" hypothesis of Jespersen further in assuming that "the motor command for the timing is the same irrespectively of the quality of the vowel, but that the execution of the command may be delayed owing to the movements to be made!". Today it must remain an open question to what extent this assumption of durational invariance is correct.

The tendency towards compensatory reorganization observed for some of the present utterances indicates an alternative possibility. It is likely that the articulatory system keeps track of the current state of articulators since evidently it takes steps to correct for the consequences that the rate and timing of mandibular movement has on the segment durations. It would make such compensations only if the
consequences of articulatory timing and rate were undesirable. Consequently the fact that the adjustments do occur may indicate that what the system is trying to achieve is relatively constant acoustic duration of each segment in a given position in spite of coarticulation and other disturbing influences.

Note in this connection the relatively constant and vowel independent duration of the final [b] in Figs. I-A-11, I-A-13, I-A-14, I-A-15. It is as if a "count-down" for the [b] begins at the moment of closure. This temporal patterning brings to mind the hypothesis of "chain-reflex" automatisms in motor timing. According to this theory (which has been criticized by Lashley (43) and Lenneberg (44)) "the performance of each element of the series provides excitation of the next" (  ).

Conclusions

(1) Evidence from articulatory modeling as well as from experimental measurements makes it appear likely that it is the dynamic behavior of the mandible that gives rise to the dependence of acoustic vowel duration on the degree of vowel opening.

(2) The universality of this dependence is thus contingent upon the extent to which the timing and rate of mandible movement observed in the present investigation for Swedish are typical also of other languages.

(3) The results suggest also that the motor control of open and close vowel duration may be characterized by less durational variability than their acoustic representations.

Summary

In spite of repeated attacks on the problem, the facts of acoustic vowel duration are still not very well understood. Nor is it known whether the physiological processes that underlie the variations of vowel duration are language-dependent or universal. Attempts to explain these facts, if successful, are likely to entail a better understanding of the motor organization of the syllable.
Vowel duration and mandible position

The point of departure of the present study is the frequently observed correlation between "tongue height" and acoustic vowel duration: The higher the tongue the shorter the vowel. We begin this paper by examining the open-close dimension of vowels, not in terms of tongue height, but in terms of jaw opening. Good agreement with the traditional vowel quadrilateral is obtained when the position of the mandible in long and short Swedish vowels is plotted against the classes of front-rounded, front-unrounded, and back-rounded vowels (Fig. I-A-1). When vowel duration in the context of bilabial stops is plotted against mandibular position a positive correlation can be observed (Fig. I-A-2): The lower the jaw the longer the vowel.

Dynamic model of articulatory coordination

A dynamic model of lip and mandible coordination is constructed that gives an articulatory interpretation of the dependence of vowel duration on jaw position. The model in which the lips and the mandible are represented by damped spring-mass systems permits the derivation of the course of the midsagittal separation of the lips, \( s(t) \) for \([bVb]\) syllables. In terms of this parameter vowel duration is defined as the interval during which \( s > 0 \) and can be determined from

\[
s(t) = x_u(t) - x_\ell(t) - x_j(t) - x_0
\]

This equation is by definition valid only when \( x_u(t) - x_\ell(t) - x_j(t) - x_0 \geq 0 \). For \( x_u(t) - x_\ell(t) - x_j(t) < 0, s(t) = 0 \). The term \( x_u \) is the labiomuscular displacement of a point on the upper lip, \( x_\ell \) is the labiomuscular displacement of a point on the lower lip and \( x_j \) stands for the jaw-dependent depression of the lower lip. The term \( x_0 \) is the criterion of closure onset or release. This model is used to compute the hypothetical course of \( s(t) \) in \([bVb]\) syllables to which we assign various degrees of vowel or jaw opening (Figs. I-A-3 - I-A-6). The model predicts that vowel duration should increase as a function of the mandibular excursion for the vowel provided that certain temporal relations hold between \( x_j \) on the one hand and \( x_u \) and \( x_\ell \) on the other: The extent of mandibular movement must be larger for open vowels even during the occlusion of the initial consonant and/or the
mandibular off-glide movement from the vowel into the final consonant must progress so slowly that the contact between the articulators for the consonant is delayed in the context of the open vowel. Moreover, it is shown that the durational variance of open and close vowels is not necessarily a characteristic of the input control signals to the model. It is compatible with invariant vowel duration at the level of control and can be generated simply as a result of the superposition of jaw movement on the opening and closing gestures of the lips.

**Method.** To expose the assumptions on which the predictions of the model had to be based data on lip and jaw movement were collected. With the procedure used the lip and the jaw could be recorded continuously and in synchrony with the speech signal. The method involved the use of cylindrical miniature lamps which were attached to the talker's lips and to a special device indicating the position of the mandible. The movement of the lamps during speech was registered by an oscilloscope camera in which the film ran perpendicular to the movement dimension under analysis. In the experiments performed measurement was made of the vertical movements of lamps on the lips and the jaw device which were aligned midsagittally.

**Results and interpretations.** Ten repetitions of four Swedish test words: [I' b:a:bI], [I' b:i:bI], [I' babbI], and [I' bibbI] were analyzed for three talkers in terms of the parameters of the model. It was found that:

1. the timing of the jaw-dependent component, in relation to the labial components and , indicates that the coordination of labial and mandibular gestures is characterized by coarticulation. The mandible begins its opening movement for the stressed vowel in [I' bV(:)b(b)I] while the lips are still in position for the [b]. Similarly, it completes its closing movement for the second [b] after the labial closure for this consonant has been attained. These effects are present for all the talkers.

2. Comparing the words with open and close vowel pairs it is seen that for some of the utterance pairs, there is reorganization of the "labio-muscular" components of lip separation . The effect of these compensatory adjustments is to prevent the open vowels from becoming even longer.
(3) $x_0$ is in general larger at the release than at the closure onset of a given [b]. This is assumed to be due to the presence of adhesion effects. It is larger also for [a:] than for [i:] at the closure of a following [b]. This is tentatively attributed to the differences in labialization between [a:] (rounded) and [i:] (spread).

References

(1) Malmberg, B.: Die Quantität als phonetisch-phonologischer Begriff (Lund/Leipzig 1944).


(3) It is known that the lowering of the mandible involves both rotation and translation (4, 5). The degree of jaw opening can be defined in various ways. Posselt (3) classifies mandibular movements in speech as "habitual (automatic) opening and closing movements". Although repetitions of these movements "do not always coincide exactly they have a fairly characteristic main course..." (Posselt, p. 40). In Fig. I-A-1 the depression of the mandible is defined as the vertical projection of a distance along this characteristic path relative to occlusion. See also below under section on "Experimental Method and Procedure".


(6) For further details on the mathematics and physics of the argument of this section see e.g., (7).


(18) Anatomical reference line running between a point on the eye orbit (orbitale) and the most lateral point in the roof of the bony external auditory meatus (porion).


(26) Ohman, S.: "Peripheral Motor Commands in Labial Articulation", article I,B in this issue of STL-QPSR.


(34) Einarsson, S.: Beiträge zur Phonetik der isländischen Sprache (Oslo 1927).


Acknowledgments

I want to express my indebtedness to G. Fant, Eva Gårding, L. Gårding, K. Hillerström, J. Lindqvist, B. Malmberg, F. Strenger and S. Öhman for valuable advice and criticism. I am grateful also to Mrs. S. Felicetti for expert assistance in typing and editing the manuscript.