DOCUMENT RESUME

ED 025 182 48 FL 001 166

By- Delattre, Pierre
The General Phonetic Characteristics of Languages. Final Report- 1967- 1968.
California Univ., Santa Barbara.
Spons Agency- Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No- BR-6- 1990
Pub Date 30 Sep 68
Contract- OEC-4-7-061990-0176
Note- 165p.
EDRS Price MF-$0.75 HC-$8.35
Descriptors- Acoustic Phonetics, *Applied Linguistics, *Comparative Analysis, Consonants, English, French, German, Graphs, *Language Research, Linguistics, *Linguistic Theory, *Phonetic Analysis, Phonetics, Phonological Units, Pronunciation, Spanish, Spectrograms, Tables (Data), Vowels

In this final stage of a series of three linguistic studies conducted at the University of California, Santa Barbara, four topics are presented. The longest is a study of consonant gemination in German, Spanish, French, and American English from acoustic, perceptual, and radiographic points of view. Pharyngeal features are studied in the consonants of the same four languages. Vowel radiography and its acoustic correlations are discussed, and a cross-language study is made of the l/r distinction. (AF)

RESEARCH CONTRACT: OEC-4-7-061990-0176
with the U.S. Office of Education
FINAL REPORT - 1967-1968

THE GENERAL PHONETIC CHARACTERISTICS OF LANGUAGES

Pierre Delattre

Project Director

University of California
Santa Barbara

October 1, 1967 to September 30, 1968

The research reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education.

TABLE OF CONTENTS

INTRODUCTION

THE PROBLEM

In the teaching of foreign languages to American students one of the major problems has always been that of acquiring a satisfactory pronunciation. Language Institutes have had more difficulty in this respect than in any other. With the present emphasis on "speaking," this problem has taken on more importance every year. In order to make an effective use of the phonemic system of a second language, one must develop good articulatory habits. Improvements in the teaching of pronunciation have been hampered by an insufficient knowledge of the segmental and prosodic features of foreign languages. Problems of interference are partly due to a lack of really objective data on the phonetic features of the first language as well as of the second -- data which would allow phonetic contrasting of the two languages in a truthful and realistic manner.

OBJECTIVES

The long range objectives of this project are the instrumental analysis and detailed description of the phonetic features of American English and of the foreign languages that are commonly taught in the United States. The foreign languages toward which our main attention is turned at present are German, Spanish, and French. Results of our investigations are to appear in article or book form.

Exploratory research has led us to divide our investigation into 40 sections -- 11 prosodic features, 13 vocalic features, and 16 consonant features. As a result, we are comparing English to German, English to Spanish, English to French, each under the 40 following headings:

Prosodic Features: 1. Declarative Intonation. 2. Non-Declarative Intonation. 3. Place of Logical Stress in the Word. 4. Place of Logical Stress in the Sense Group. 5. Nature of Logical Stress. 6. Place of Emphatic Stress. 7. Nature of Emphatic Stress. 8. Variations in Syllable Weight. 9. Internal Juncture and Syllabication. 10. Syllable Type. 11. Tension.

Vocalic Features: 12. Articulatory Description. 13. Acoustic Description. 14. New Vowel Sounds for the English Speaker. 15. Distribution (Positional and Allophonic). 16. Frequency of Occurrence. 17. Duration System. 18. Neutral Position. 19. Loss of Color. 20. Effect of Consonant Anticipation on Vowels. 21. Diphthongization. 22. Effect of Syllable Type on Vowel Color. 23. Attack and Release. 24. Nasality.

Consonantal Features: 25. Articulatory Description. 26. Acoustic Description. 27. New Consonant Sounds for the English Speaker. 28. Distribution (Positional and Allophonic). 29. Frequency of Occurrence. 30. Duration System. 31. Neutral Position. 32. Consonantal Weakening.

33. Effect of Vowel Anticipation on Consonants. 34. Speed of Articulation. 35. Tongue Fronting. 36. Aspiration. 37. Affrication. 38. Palatalization. 39. Final Release. 40. Voicing.

PROCEDURES

In order to complete these investigations with a high degree of objectivity, we have developed a three-way instrumental technique of research based on the successful design and construction of special instrumentation.

1) This three-way technique generally begins with the spectrographic analysis of utterances that have been composed and recorded for a special purpose. The comparison of spectrograms of English with corresponding ones of German, Spanish or French leads to making some hypotheses to the acoustic differences between English and the other languages, regarding certain phonetic characteristics.

2) Then, the hypotheses are verified or refuted by means of spectrographic synthesis. Spectrographic patterns of the contrastive utterances are painted and transferred into sound by a speech synthesizer. It is thus possible to judge by ear to what extent the assumed acoustic differences produce the appropriate auditory differences.

3) Finally, motion picture x-rays of the utterances are made and studied frame by frame by means of special projectors, to discover the articulatory features that correlate with the acoustic ones found by spectrographic analysis and synthesis.

4) As a complement to this instrumental research, phonetic features of English and foreign languages are investigated by statistical analysis, related to such features as phoneme frequency, phoneme distribution, syllable types, etc.

RELATED RESEARCH

Our investigation of the phonetic characteristics of languages is related to research in the general field of Applied Linguistics in that it will contribute to our conclusive knowledge of English and foreign languages and will make it possible to improve their teaching.

It is also related to research in Methodology, seeing that experiments aimed at obtaining better results in language teaching will use our data.

Finally, it is related to research in General Linguistics because our acoustic and articulatory studies are closely connected with the determination of the "distinctive features" of phonemes and prosodemes.

RESULTS

During this year of research under contract with the Office of Education we have completed four studies, the texts of which follow.

A CROSS-LANGUAGE STUDY OF THE j/i DISTINCTION

This is a "contrastive" study in the cross-language sense of the word. The semi-vowel /j/ of yes [jɛs] poses problems of phonetic interference to the language teacher when it occurs after a consonant. Pronounced with French articulatory habits, for instance, the English word radio can be unintelligible to an American ear, and, conversely, the French word radio can be unintelligible to a French ear when pronounced by an American student.

This phonemic status of /j/ in the two languages is comparable. In English as in French, post-consonantal /j/ contrasts with other consonants (E., Cue /kju/, Clue /klu/, Crew /kru/, F., Quiet /kjɛ/, Clef /klɛ/, Craie /krɛ/), with zero (E., Cue /kju/, Coo /ku/, F., Quiet /kjɛ/, Quai /kɛ/), and with /i/ (E., It's Lilian her cousin /ɪts lɪljən hɚ kʌzən/, It's Lily an(d) her cousin /ɪts lɪli ən hɚ kʌzən/, F., Si Julia paraît /si ʒylja parɛ/, Si Julie apparaît /si ʒyli aparɛ/).

The purpose of this study is to relate the articulatory and auditory habits, for the pronunciation of post-consonantal /j/, to objective data obtained by sound spectrography, by cineradiography, and by the techniques of artificial-speech synthesis.

1. SPECTROGRAPHIC ANALYSIS

In order to contrast the phonetic behavior of English with that of French /j/, we shall first examine the spectrograms of the two sentences already mentioned: <u>The pianist from Vienna plays on the radio</u> and <u>Le pianiste viennois joue à la radio.</u> In the English syllables <u>pia-</u> and <u>Vie-</u>, the /j/ stands before a stressed vowel; in the syllable <u>-dio</u>, it stands before an unstressed one. In French the situation is reversed -- in the syllables <u>pia-</u> and <u>vie-</u>, the /j/ stands before an unstressed vowel; whereas in the syllable <u>-dio</u> it stands before a stressed one. Both types of stress-conditions, therefore, are included in our sentences. However, differences caused by the place of the stress will not prove to be significant.

In Fig. 1, the spectrographic patterns of these two sentences are presented in two ways. First, the spectrogram of human speech, spoken at a normal syllabic rate without emphasis, is given. On such a spectrogram only the most obvious phases of articulation appear. Relevant details are left out because of inadequate resolution of the low intensities. To provide all the acoustic details that are masked on the ordinary spectrogram, a synthetic pattern of the sentence is drawn under each of the human spectrograms. This hand-painted spectrogram, when passed through a speech synthesizer, is transformed into the sounds of the desired sentence in a very intelligible manner. It serves here as a practical reference to the acoustic cues that are necessary for the sentence to be clearly understood, even in slow motion.

To learn a little how to read the spectrograms of Fig. 1, let us look, in each case, at the synthesis under them. These synthetic patterns show, at all times, three formants as three black lines of variable thickness, undulating up and down from left to right. On the spectrograms, these formants are seen as

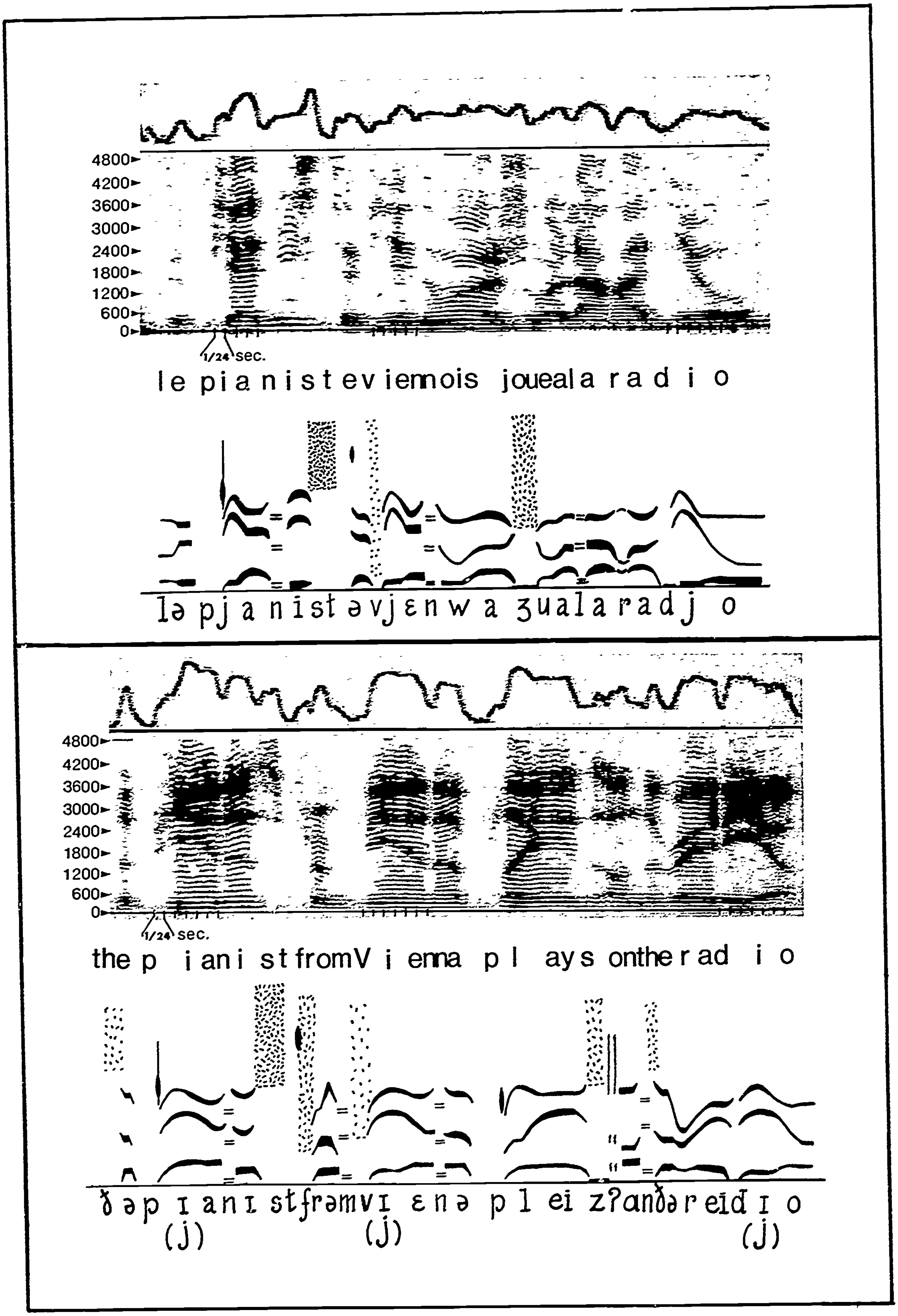


Figure 1. Spectrograms and synthetic-speech patterns for French and English /j/.

the three first (from the bottom) darker concentrations of energy or darker groups of harmonics. (Every thin horizontal line is a harmonic.) Formants reflect the resonance notes of mouth cavities. As these cavities change their volume and their shape during the speech process, the formants change their frequency and intensity as a factor of time. The intensity of formants is shown by the degree of darkness; their frequency in cps (cycles per second) can be estimated by the frequency scale on the left, and the time is given in units of 1/24th of a second along the zero line (bottom of the spectrogram) in order to correspond to the frames of x-ray films which are taken at 24 frames per second and which will be seen later. 1/24th of a second equals about 4 cs.

A fourth (and at times even a fifth) formant appears in the spectrograms. It is not included in the synthesis because it has no linguistic function (it does not affect meaning); rather it is related to voice quality and describes the speaker himself, not what he said.

Above all the formants, on the spectrograms (only), is a line showing the variations of overall amplitude (the sum of the amplitude of all the harmonics).

After the article (le in French, the in English), an interruption is seen which corresponds to the lip closure for the /p/. Then the first, second, and third formants rise sharply. These rises (called transitions) are necessary for the production of labial stops. (If the second formant were falling instead of rising, for instance, the sound would be that of a velar /k/, and if it were moderately rising, instead of sharply, the sound would be that of dental, or alveolar, /t/.) A vertical line immediately preceding the rising formant-transitions represents the burst of noise of the /p/ explosion which occurs when the lips separate abruptly. This burst is visible on the French spectrogram but not on the American one.

The /p/ formant-transitions, on both spectrograms, start without voicing (the vocal cords have not yet started vibrating); they are composed of noise rather than harmonics. In English the voiceless transitions occur during "aspiration." They last for nearly 10 cs until the formants have reached about the level of a low [i]; the second formant lingers at that level, then falls toward an [æ] level just before the brief /n/ closure (hold). (The first formant of the /a/ fades out as a sign of nasalization, the velum having lowered by anticipation of the following /n/.)

In French, the 7 or 8 centiseconds of voiceless sound which follow the /p/ explosion include most of the /j/, which has become voiceless by assimilation to the voiceless /p/. (The main law of assimilation states that the stronger of the two consonants in contact dominates -- the /p/ being stronger passes on its voicelessness to the /j/ and makes of it a sort of [ç] as ch in German ich.) During this short interval of time, considerable acoustic change takes place, according to the synthetic pattern. The second- and third-formant transitions rise abruptly to the /i/ level and fall abruptly to the /a/ level, which they have nearly reached when the vocal cords start vibrating (that is, when the harmonic lines reappear).

Let us now compare the English and French /j/ as objectively as we can. Four parameters of the /j/ can be measured.

(a) Durations of the English and the French /j/. From the explosion of the /p/ to the hold of the /n/, the English /pja/ measures about 24 cs, whereas the French /pja/ measures only about 16 cs for a ratio of 3 to 1. Out of that length of 16 cs for the French /j/, about 4 cs go to the rising /p/-transitions and 8 cs to the /a/ steady-state. This leaves 4 cs for the French /j/. Out of the 24 cs of the English /j/, about 8 cs go to the rising /p/-transitions and at most 6 cs to the /a/. This leaves at least 10 cs for the English /j/, perhaps much more, depending on the position of the falling transitions which are responsible for the perception of the /a/. In short, the English /j/ is at least 6 cs longer than the French /j/.

(b) <u>The voiced portion</u> of the /j/. In English, voicing (noticeable when harmonics become visible in the formants) begins immediately after the second-formant rise for /p/; it includes all the slow-falling curve of /j/. The English /j/ must, therefore, be considered entirely voiced. In French, looking again at the second-formant transition, voicing begins at the end of the falling curve of /j/, just before the /a/ begins. This leaves at least one-half, perhaps three-quarters, of the /j/ as voiceless. In short, a larger proportion of the /j/ is voiced in English than in French.

(c) <u>The second-formant frequency</u>. At its highest point, the second-formant frequency of the English /j/ is at about 2100 cps, which is equivalent to the frequency of an /ɪ/ as in <u>bit</u>; the second-formant frequency of the French /j/ is at about 2500 cps, which is equivalent to the frequency of a high /i/ as in <u>si</u>. In short, the second-formant starting level is considerably lower for English /j/ than for French /j/.

(d) The speed of formant transitions. The rate of change of the second-formant frequency downward move toward /a/ is about 400 cps in 10 cs for the English /j/, which falls from 2100 cps to 1700 cps; it is about 900 cps in 4 cs for the French /j/, which falls from 2500 cps to 1600 cps. In short, the speed of formant transitions is much slower for the English /j/ than for the French /j/.

To summarize the measurable differences between the English and the French /j/ in pianist, it can be said that the English /j/ is more voiced, longer, lower in formant-two frequency, and slower in transition-two rate of change. These factors all seem to mean, in subjective terms, that the /j/ is more vocalic or less consonantal. The vocalic feature is especially applicable to the last parameter, the rate of change, which is the one that best distinguishes consonants from vowels -- vowels are essentially perceived by a steady state in the frequency of formants; consonants are essentially perceived by a frequency change in the formants (hence the name: formant transitions). The faster the change, the more consonantal the sound, and vice-versa.

Let us now compare the /j/ phonemes in English Vienna and French viennois. Here the /j/ is preceded by a voiced consonant, and there can be no unvoicing assimilation. (Even though the fundamental, or first harmonic, does not show on the spectrograms of either the French or the English /v/, we must assume that it is voiced throughout, that the fundamental waves are not strong enough to get through the /v/ constriction.) So, no difference appears with respect to voicing. But there is a very clear difference in the /j/ intensities (amplitude display line above the spectrogram).

For English Vienna, after the amplitude depression for /mv/, the line rises sharply to the vowel level of /ɛ/ and remains at that same high level for the whole duration of the /j/. In short, the sequence /vjɛ/ shows only two amplitude levels, one for /v/, another for /jɛ/. In the case of French viennois, the sequence /vjɛ/ clearly shows three levels: a low one for /v/ (shorter than in the English utterance because it is not preceded by /m/); a mid level for /j/, and a high level for /ɛ/. The English /j/ in Vienna is, therefore, more intense (in subjective terms, louder) than the French /j/ in viennois. In articulatory terms, less intensity should correspond to a narrower constriction.

In addition to intensity features, the differences already observed between English pianist and French pianiste are also visible in Vienna vs. viennois: the English /j/ is longer, lower in formant-two frequency and slower in transition-two rate of change. To see all that, one must look at the synthesis patterns rather than the spectrogram patterns, which have poor intensity resolution of the fast and low formant changes of /j/.

A comparison of the syllables /djo/, in English radio and French radio, shows all the divergences observed in /vj/. The amplitude for the French /djo/ does not rise in three separate steps as clearly as for /vjɛ/, however, but it does rise much more gradually than in the English /djo/ from the /d/ depression to the /o/ maximum. In English, after the /d/ depression, the /j/ amplitude rises sharply to the /o/ level. (Actually, in the English utterance, the /o/ is even less intense than the /j/, perhaps because it is more unstressed as the end of the word is nearing.)

To summarize, spectrograms, with the help of synthesis, show objectively five acoustic differences between the English and the French /j/ in the sentences: The pianist from Vienna plays on the radio and Le pianiste viennois joue à la radio. The English /j/ is (a) more voiced, (b) more intense, (c) longer, (d) lower in formant-two frequency, and (e) slower in transition-two rate of change.

In articulatory terms, this should mean (a) more vibrations of the vocal cords, (b) a wider palatal constriction, (c) more time given to the articulation, (d) a larger front cavity (tongue lower and less fronted), and (e) a quicker articulatory shift to the next vowel. (It is better not to say a quicker 'opening' motion even though this is generally the case, because in syllables like /jn/ the opening remains nearly constant, both sounds being close.)

In subjective terms, the English /j/ might well be called more "vocalic." Each one of the five traits just mentioned is a vocalic trait, especially the last one -- a slow rate of change in the formants.

Before leaving these spectrograms of English and French utterances, the reader will not fail to notice some other differences not related to the /j/ sounds. For instance, the hold (closure) time of the English /n/'s and the English /d/ are much shorter than the French ones. These American dentals, in unemphatic articulation, look almost like Spanish /r/ flaps. Another noticeable difference is the clear occurrence of an [ə] between pianiste and viennois in French. No trace of such an obtrusive sound is visible in English after pianist even though the conditions required in French for its occurrence are equally present -- between consonants after two consonants or more.

2. THE X-RAY ANALYSIS

Fig. 2 presents a sequence of cineradiographic frames for the articulation of the words pianist (left) and pianiste (right) by native speakers of English and French, respectively. These sequences were taken in our own x-ray studio at 24 frames per second, with simultaneous sound, and can be studied by means of special projectors, at normal speed and in slow motion while listening to the sound.

Next to the x-ray sequence, the spectrograms of the same words indicate by means of arrows the acoustical point in time which corresponds to each articulatory frame. These arrows are moved forward by about 4 centiseconds at each frame.

The x-ray films include a sufficient portion of the head and neck to show the whole vocal tract, from the lips to the vocal cords, that is, the complete resonating system of mouth cavities. In Fig. 2, the level of the vocal cords is indicated by the horizontal line at the bottom of the pharyngeal cavity. The vertical line limiting each image to the right is the pharyngeal wall, ending in the upper-right corner at the rhino-pharyngeal cavity. In the same corner is the velum, or soft palate, which can either shut the velic corridor and prevent communication between the mouth cavities and the nasal cavities, as in frames 1, 2, and 3 of Fig. 2, or open that corridor to let the nasal cavities combine their resonance with that of the mouth cavities,

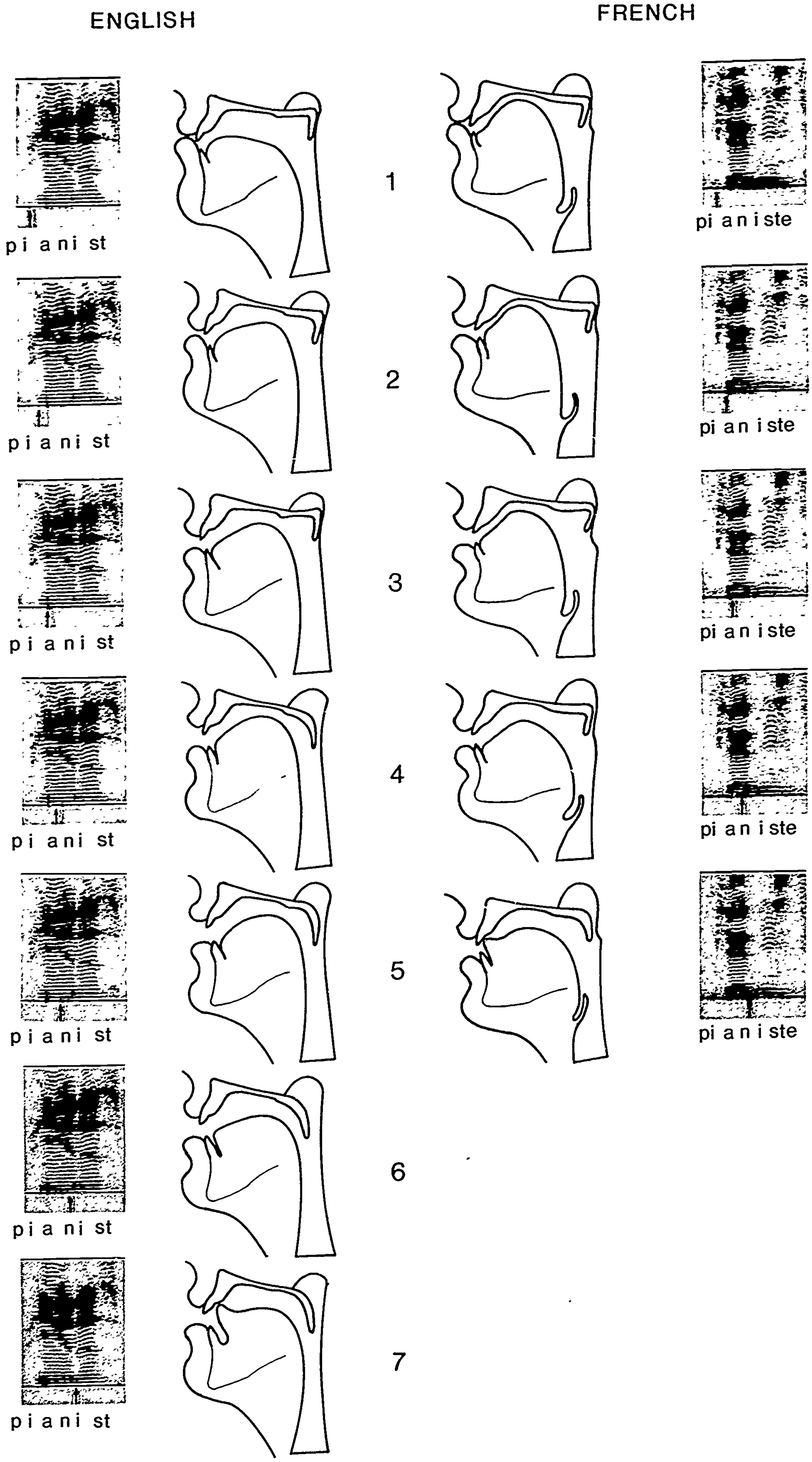


Figure 2. X-ray frames for the comparison of English and French /j/.

as in frames 4 to 7 of Fig. 2. Continuing from right to left, we see the velum (soft palate) attached to the bone of the hard palate. The palatal bone ends with the alveolar ridge, the upper incisors, and the upper lip. The lower lip and the chin surround the lower incisors and the lower-jaw bone. Finally the tongue, in the center, appears as a mid-line profile.

The x-ray sequences of /j/, selected from the words pianist and pianiste (Fig. 2), go from the last frame of /p/ to the first frame of /n/. Between /p/ and /n/ the English /j/ occupies frames 2 to 6 and the French /j/ frames 2 to 4.

The first thing to notice is that the English /j/ is longer than the French /j/ -- 5 frames here and 3 there.

Let us compare the English film and the French film frame by frame.

Frame 1. At the last contact of the lips before they separate abruptly for the /p/ explosion, the tongue positions of English and French are sharply different and illustrate remarkably the behavior of the two languages with respect to vowel anticipation. At frame 1, the French speaker has practically anticipated the tongue position for [ɪ] or [j] -- the tongue dorsum is high and fronted enough for an acceptable [ɪ] but the tongue root is still not far enough from the pharyngeal wall. In fact, the tongue is only one frame away from the highest position it is going to assume. At the same frame 1, the American speaker's tongue is three frames away from the highest position it is going to assume. As the lips separate, the tongue shows no anticipation at all of the coming /j/, occupied as it is in producing a labial consonant by lowering the tongue.

At frame 2, the French speaker's tongue is well fronted, as for [j] rather than [i], not only by its approximation to the palate, in front, but also by its distance from the pharyngeal wall. The narrowness of the palatal constriction partly explains the production of the noise which can be seen on the spectrogram. At the same moment (frame 2) the American speaker's tongue is leisurely moving toward a high front position -- the tongue dorsum is rising and the tongue root is widening its distance from the pharyngeal wall. Furthermore, the opening of the lips is taking place more slowly than in the corresponding French frames.

At frame 3, the French speaker has just started his tongue motion toward the /a/ -- the lips and jaws have widened their separation, and the tongue has been slightly lowered at the palate and drawn back toward the pharyngeal wall at the root. At frame 3, the English speaker's tongue is leisurely continuing its rise toward a high-fronted position, and his lips and jaws continue separating.

At frame 4, the French speaker's tongue has reached the closest position to an [a] that it will take on the film (it might have gone farther between frames). The main requirement for the articulation of an [a] is that the tongue root make a constriction along the lower part of the pharyngeal wall; the second is that the tongue be considerably lowered and the jaws well apart. Both requirements are observed in frame 4, but minimally. This poor articulatory realization of an [a] is perhaps due to two combined factors -- the unstressed position of /a/ and the anticipation of the /n/. At frame 4 of the

English articulatory sequence, while the French tongue has already reached the /a/ position, the American tongue is only reaching the high-fronted position from which it will move downward and backward for the final phase of the /j/ in frames 5 and 6.

The tongue position for /n/ is reached in French at frame 5 with the tip contacting the upper incisors and in English at frame 7 with the tip contacting the alveols in a typical retroflex shape.

Another difference related to the /n/ is the lowering of the velum for the production of nasality. In the English sequence the velum begins to lower at frame 4, therefore, three frames ahead of the tongue-tip contact. In the French sequence the velum starts moving away from the pharyngeal wall at frame 4, only one frame ahead of the tongue-tip contact. Consonant anticipation is, therefore, much more pronounced in English than in French.

These two sequences, then, offer a comparison of both vowel anticipation and consonant anticipation. The first frames of the sequence show a marked tendency toward vowel anticipation in French and a total lack of such tendency in English, whereas the last frames show a strong tendency toward consonant anticipation in English and only a weak one in French.

In brief, the articulatory sequences of our x-ray films confirm four of the articulatory assumptions we had made earlier on the basis of spectrograms: (a) they give no indication concerning voicing -- the vocal folds are visible on profile x-rays, but whether the cords vibrate or not cannot be detected -- but they show (b) that the constriction is wider, (c) that the whole /j/ articulation is longer, (d) that the front cavity is larger, and (e) that the articulatory movements are faster in English than in French. Furthermore, our x-ray sequences suggest that these four features are related to differences in articulatory habits of a broader nature -- the habits of vowel anticipation (weak in English, strong in French) and consonant anticipation (strong in English, weak in French).

The x-ray sequences of Vienna vs. viennois and radio vs. radio, in Figs. 3 and 4, bring nothing very new, but they confirm the differences discovered in Fig. 2 and show that they are not occasional but are characteristic of the post-consonantal /j/ articulation in English and in French.

In Fig. 3, while the upper teeth contact the lower lip for /v/, the French tongue shows much more anticipation of a high-fronted position than does the American tongue. Conversely, in anticipation of the /n/, the American velum is already withdrawing from the pharyngeal wall at frame 3, four frames ahead of the tongue-tip contact, whereas the French velum withdraws at frame 4, only one frame ahead of the tongue-tip contact.

The highest rise of the tongue in preparation for the downward motion of the /j/ is at frame 2 in French, but only at frame 3 or 4 in English. The /ɛ/ position is reached at frame 4 in French, but only at frame 6 in English. The whole articulation is longer, slower, and less constricted in English than in French.

In Fig. 4, differences of consonant anticipation show, not at the end, but at the very beginning of the English sequence,

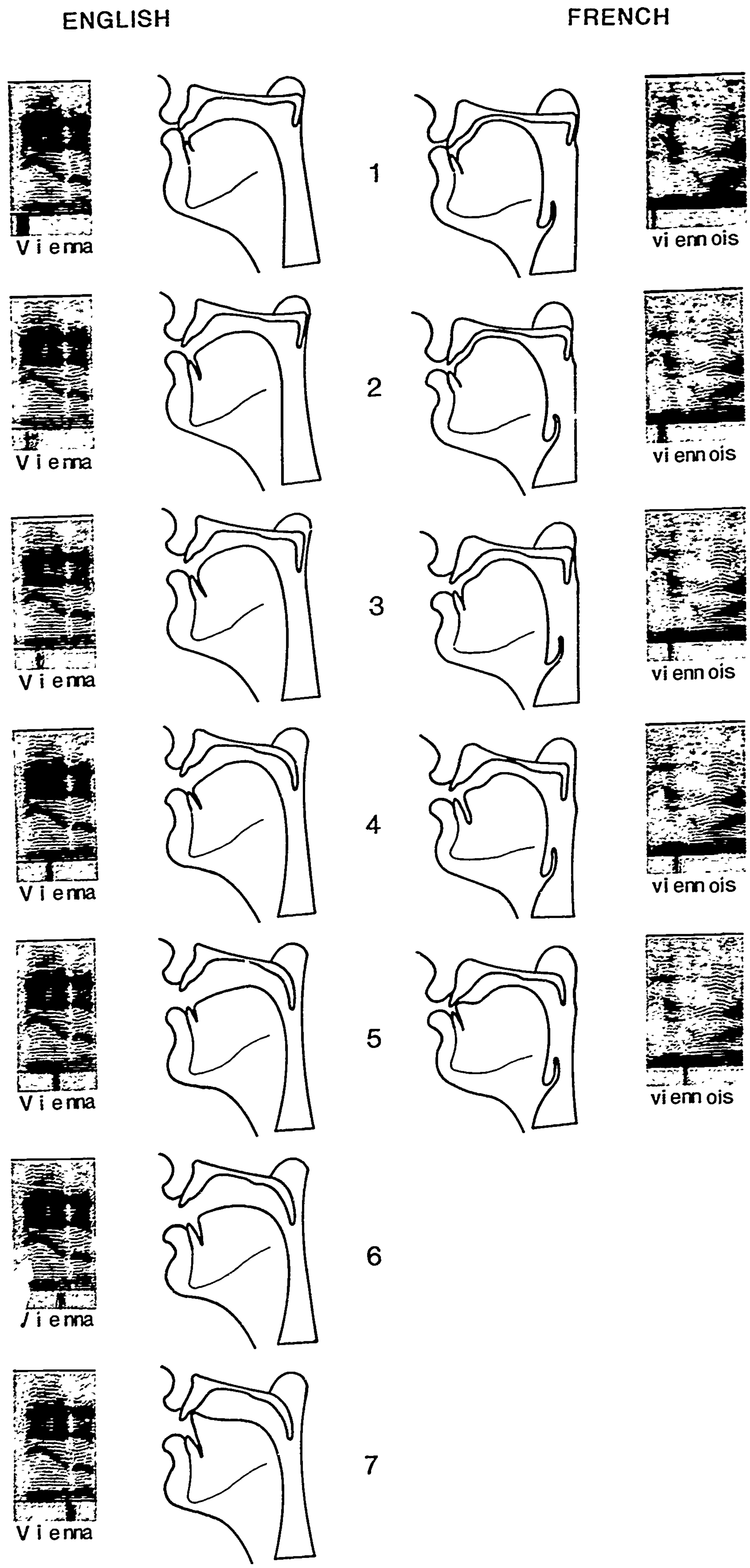


Figure 3. X-ray frames for the comparison of English and French /j/.

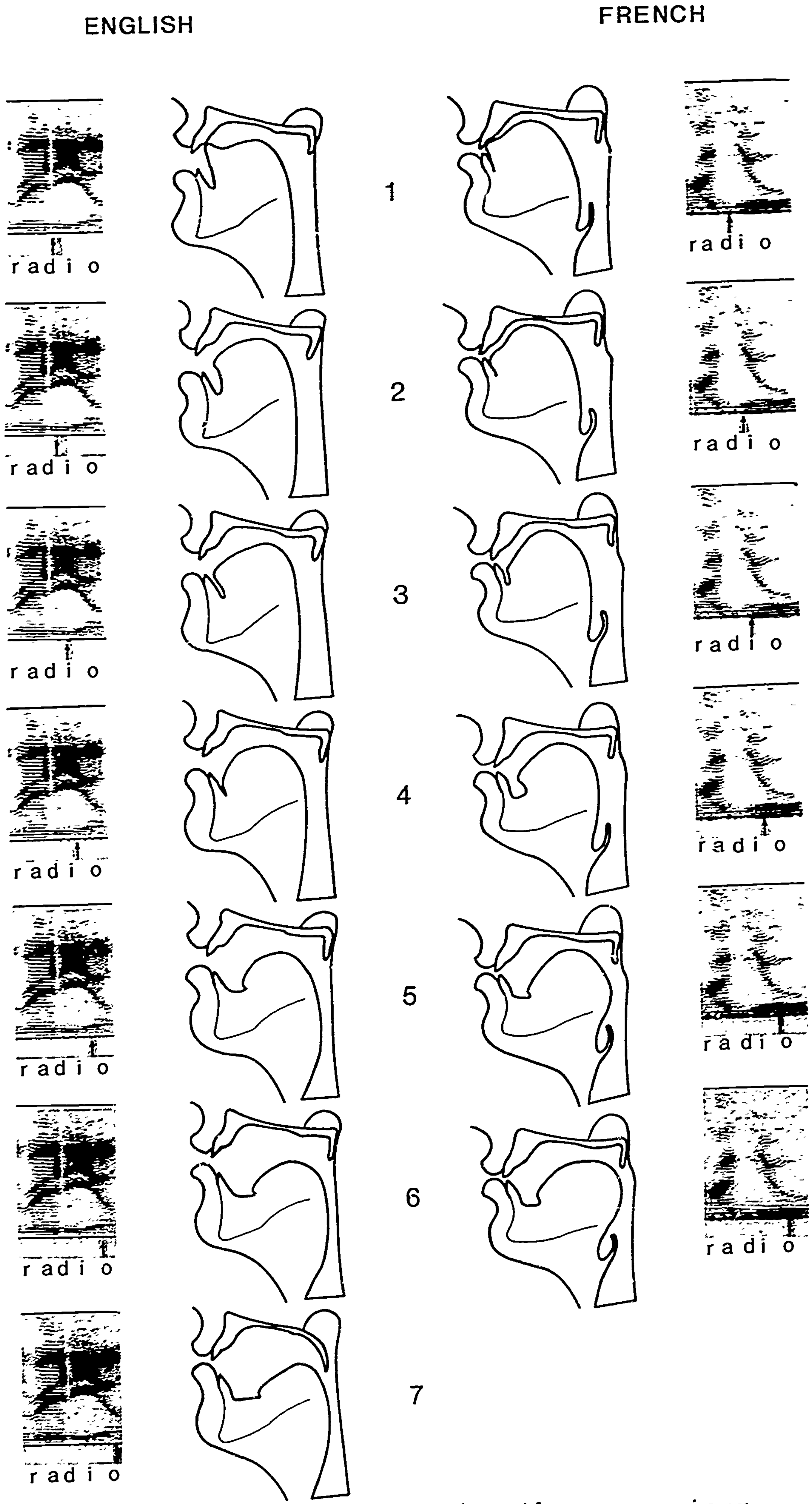


Figure 4. X-ray frames for the comparison of English and French /j/.

in frame 1, when the tongue-tip contacts the alveols while the jaws are still open for the preceding [e] of radio. In the French sequence the tongue-tip contact of /d/ does not refer back to the jaw opening of the preceding [a], but to the jaw opening of the following /j/.

Vowel anticipation can be compared, here, in the lip action for the /o/ of radio. Lip rounding and closing for the coming /o/ is practically as advanced at frame 5 of the French sequence as at frame 7 of the English sequence.

Differences in duration, in speed, and in aperture are perhaps not so pronounced in Fig. 4 as in Figs. 2 and 3; nevertheless, they are clear and convincing. The fact that, in French, the /j/ to /o/ tongue-motion takes more frames than the /j/ to /a/ motion for piano, or the /j/ to /ɛ/ motion for viennois is simply due to a difference in the articulatory distance. In English, the articulatory distance from /j/ to /o/ is less marked than in French because /j/ starts approximately at an [ɪ] level rather than an [i], and /o/ ends nearly at an [ɔ] level rather than [o] -- the lips are close as for [o], but the pharyngeal cavity barely has an [ɔ] volume as compared with the larger pharyngeal cavity of the French /o/.

3. THE PERCEPTUAL TEST

Up to this point, our comparison of French and English has been limited to the articulatory level and has relied upon the eye; we have analyzed the acoustic and articulatory specifications of speech that are made visible by the camera and the spectrograph. Our comparison must now rely upon the ear. We want to find out whether French and American ears perceive the same sounds similarly or differently when they listen to them linguistically, that is, in a meaningful context. And if differences are found, we want to know the acoustic specifications of the stimuli which incite the perception of the English phoneme and those which incite the perception of the corresponding French phoneme.

To yield clear results, this sort of comparison must be based on no more than one acoustic factor of the linguistic opposition at a time. Here we have chosen to test the factor of duration alone, the /j/ of words like radio having regularly shown a divergence of length in the acoustic and articulatory analyses of the first and second part of this investigation. For this purpose we have used the linguistic alternation i/j in the following utterances:

	It's Lily an(d) her cousin / /ɪts lɪli ən hɚ kʌzn̩/	(1)
vs.	It's Lilian her cousin / /ɪts lɪljən hɚ kʌzn̩/	(2)
and	Si Julie apparaît / /si ʒyli aparɛ/	(3)
vs.	Si Julia paraît / /si ʒylja parɛ/	(4)

These utterances were synthesized in such a way that the distinction of meaning between utterances (1) and (2) as well as between utterances (3) and (4) would depend exclusively upon the duration of the steady state in the i/j formants. Figs. 5

and 6 present the synthetic patterns which, in spectrographic form, produce the sounds of utterances 1, 2, 3, and 4, when passed through a speech synthesizer of the pattern playback type.

For those who might be interested in the synthesis of the four utterances, we show below how simply it can be explained. This explanation is put in brackets to indicate that it may be by-passed.

[The English utterance. Each vowel is represented acoustically by three broad horizontal lines. These lines are called "formants" because they reflect the varying volumes and shapes of the mouth cavities and the frequencies at which they resonate. Most of the time, only the first and second formants (the two lowest) are relevant for the distinction among vowels -- the third formant is generally in the vicinity of 2500 cps; it is markedly higher only for [i] of which there are none in the English utterance but two in the French utterance (Si and -lie), and it is markedly lower for [ɚ] only, as in her. We can, therefore, generally limit ourselves to observing the first two formants. The /ɪ/ of It's has F_1 at 400 cps and F_2 at 2100 cps, which is normal for an /ɪ/; but the /ɪ/ of Lily has F_1 at 500 cps and F_2 at 1500 cps (which is closer to [ʌ] than to [ɪ] because it is strongly centered by the influence of the adjacent /l/'s. The final /i/ of Lily has formants between those of an [i] and those of an [ɪ]. The [ə] of unstressed an(d) has about the frequency of the first and second formants of [ʌ] in Lilian, or /ʌ/ in cousin; and so has [ɚ], but this American vowel is distinguished from [ə] and [ʌ] by F_3 which is so sharply lowered that it almost coincides with F_2.

Practically speaking, therefore, there are only three different vowels in this utterance, that of It's and -ly, that of Li-, an(d), and cou-, and that of her. As to the consonants, they are produced by the so-called Transitions (T), or rapid changes of formant frequency, which can be seen at the beginning and the end of vowels. For the alveolars /t/, /s/, /n/, /z/, T_2 points to a mid level of 1800 cps, T_3 to a high level of 2700 cps. For the velar /k/, T_2 points to a higher frequency than for alveolars and T_3 to a lower one. For the American lateral /l/, especially the post-vocalic one, both T_2 and T_3 point lower than for the alveolars /tsnz/ that are not lateral. The consonant /h/ is produced by noise (random dotting in the figure), rather than harmonics, at the levels of F_2 and F_3 of the following vowel (absence of sound at the F_1 level is essential for all friction sounds). The murmurs of /l/'s are essentially distinguished from those of /n/'s by the frequency of F_1, which is lower for /n/ than for /l/ -- 250 cps vs. 400 cps. Naturally the intensity of /n/ and /l/ murmurs is much lower than that of normal vowels; hence, the thin lines by which murmurs are represented. High random dotting represents the /s/ and /z/ friction noises. Voiced /z/ is shorter than voiceless /s/, as voiced consonants are always shorter than their voiceless counterparts.

The French utterance. Si Julie apparaît includes four different vowels which require three levels of each formant -- for F_1, the low level of /i/ and /y/, the high level of /a/, and the mid-high level of /ɛ/; for F_2, the high level of /i/, the mid-high level of /y/ and /ɛ/, and the mid level of /a/. F_3 is higher for /i/ than for the other vowels. Formant transitions show more variety of place and manner of articulation than in the preceding English utterance. For dental /s/, T_2 points to a mid frequency and T_3 to a high one. For post alveolar /ʒ/, T_2 points to a higher frequency than for /s/ and T_3 to a lower one. For labial /p/, both T_2 and T_3 point to low frequencies. The French /l/ of Julie has a much higher second-formant murmur than the American /l/'s of Lily. This is characteristic of the two languages for all /l/ sounds. The pharyngeal /r/ is distinguished from the fronted lateral /l/ by a higher F_1 and a lower F_2, with transitions joining the next vowels according to their frequency. The random noise (high dotting) of /ʒ/ is lower (articulated farther back) and shorter (voiced) than that of /s/. The formant transitions of /j/ in Julia start from the [i] levels and shift immediately toward [a] levels, instead of maintaining a high [i] position before shifting, as in Julie.]

Let us now examine the English utterances of Fig. 5.

It can easily be seen that the /i/ of Lily in the upper pattern is held for a long time. Its formant frequencies are in a steady state for 18 cs (nearly one-fifth of a second). The /j/ of Lilian, in contrast (lower pattern), is very short; its second formant has no steady state at all and is measured as zero cs. These are the two extremes between which eight other patterns were painted and synthesized with durations of the i/j varying from 2 cs to 16 cs in steps of 2 cs, as shown on Fig. 5 by dots under, or after, formant two. Thus, 10 patterns were created and transformed into sound. All their acoustic features were similar except for the length of the i/j which had durations

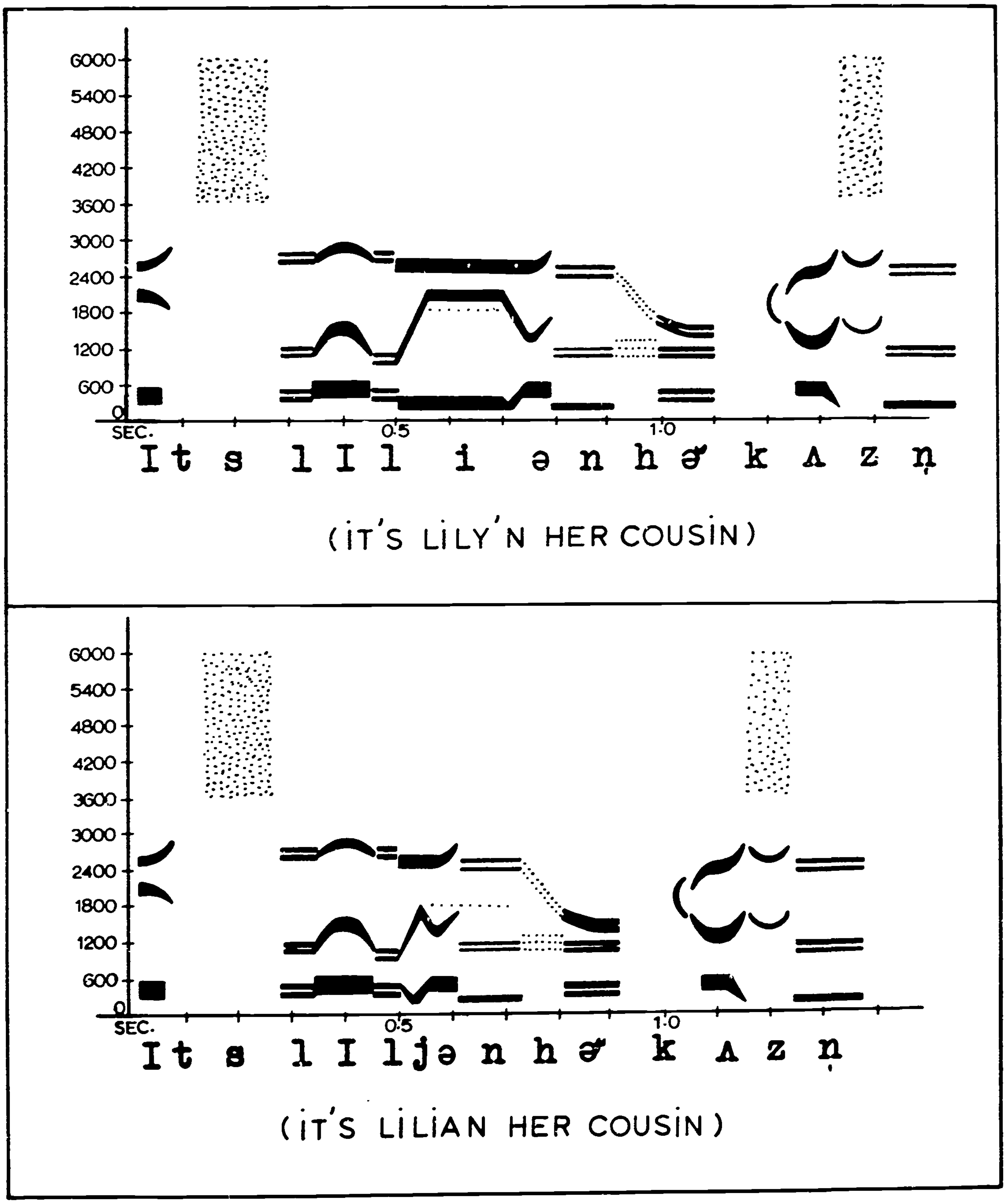


Figure 5. Synthetic-speech patterns for variation of i/j duration.

of 0, 2, 4, 6, 8, 10, 12, 14, 16, and 18 cs.

The ten patterns were recorded on magnetic tape five times each. The resulting 50 patterns were separated, mixed in random order, spliced together again, and presented for perceptual judgements by ear to 21 naive listeners -- 20 speakers of English and 1 speaker of French. The listeners were asked to mark on test sheets whether they heard each stimulus as _It's Lily_ or as _It's Lilian_.

The French utterances of Fig. 6 were treated in exactly the same manner. It can be seen that the i/j is much longer in _Julie_ than in _Julia_. The dots under formant-two of _Julie_, or after formant-two of _Julia_, indicate the ten different lengths given to the i/j alternation. A similar test of 50 items was prepared by mixing the stimuli in random order and was presented for perceptual judgements by ear to 21 naive listeners -- 20 speakers of French and 1 of English. The listeners were asked to mark on test sheets whether they heard each stimulus as _Si Julie_ or as _Si Julia_.

Results and discussion. The results of the perceptual test are given in Fig. 7 in the frame of coordinates which indicate the steady-state duration of the i/j formants on the ordinate and the number of /j/ judgements (that is, the number of times _Julia_ rather than _Julie_, or _Lilian_ rather than _Lily_, was heard) on the abscissa. There are four test results, two for French subjects listening in one case for French utterances (upper left), in the other case for English utterances (lower left); and two for American subjects listening in one case for English utterances (upper right), in the other for French utterances (lower right).

(a) We had asked one French subject to judge the English

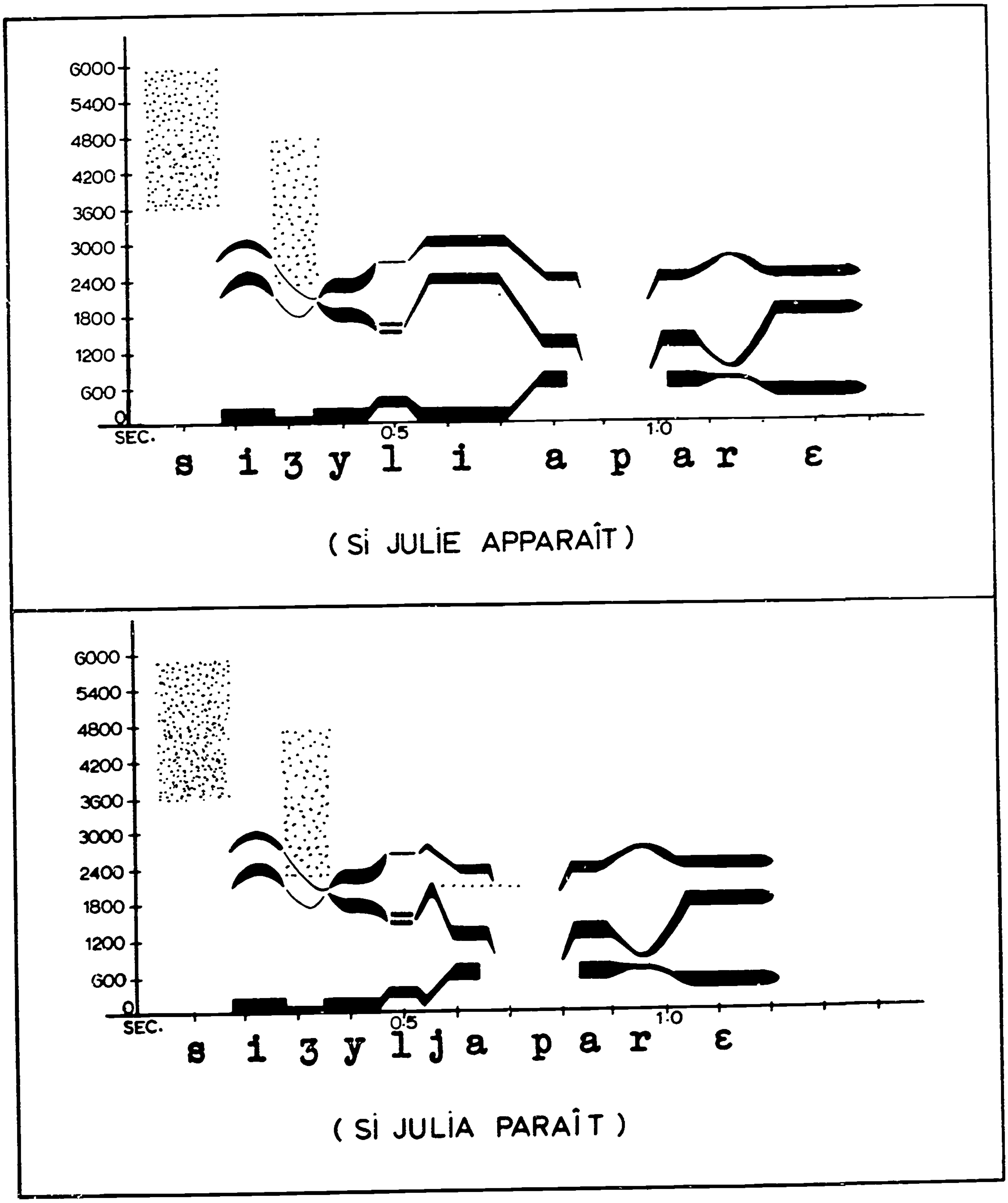
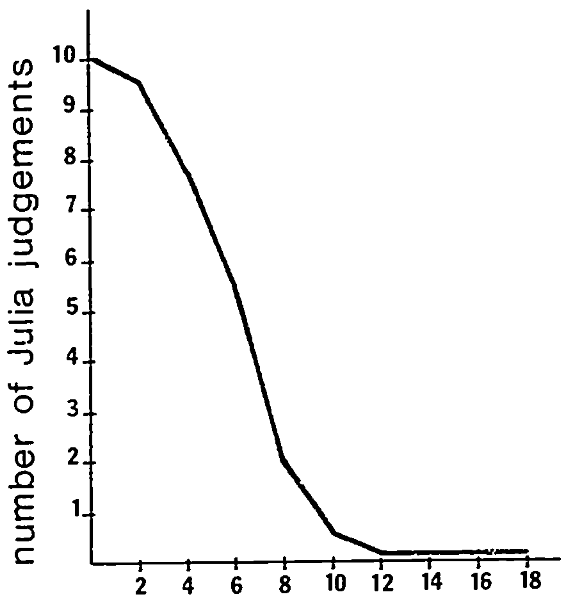
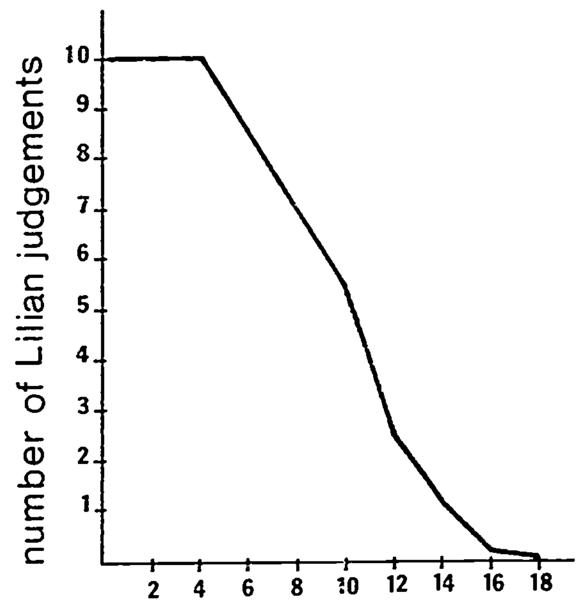


Figure 6. Synthetic-speech patterns for variations of i/j duration.



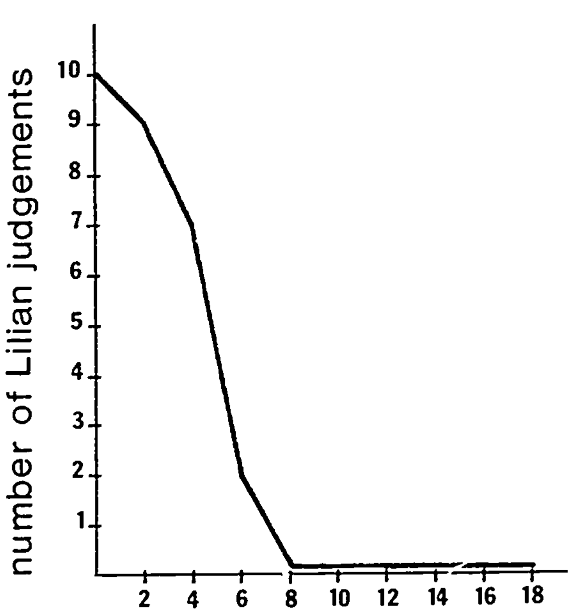
duration of i vowel in Julia-Julie

judgement by 20 French subjects



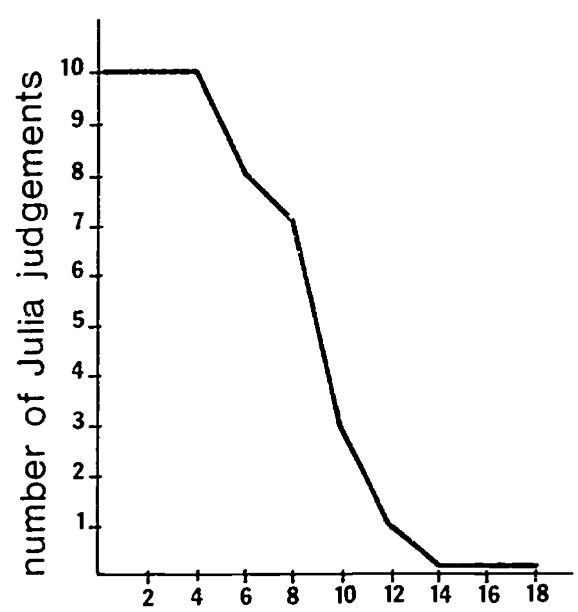
duration of i vowel in Lilian-Lily

judgement by 2C American subjects



duration of i vowel in Lilian-Lily

judgement by 1 French subject



duration of i vowel in Julia-Julie

judgement by 1 American subject

Figure 7. Comparative identification of /i/ and /j/ by French and American listeners.

utterances as well as the French ones, and one American subject to judge the French utterances as well as the English in order to verify the validity of our testing technique. This technique being based on presenting English language stimuli to American listeners and French language stimuli to French listeners, the comparison of American results with French results might perhaps not be acceptable, even though the physical conditions of the j/i alternation were extremely similar. The two lower curves of Fig. 7 show that our doubts were unjustified -- the results of the French subject judging English utterances (lower left) are not significantly different from those of all French subjects judging French utterances; and the results of the American subject judging French utterances (lower right) are not significantly different from those of all American subjects judging English utterances. It can be assumed, therefore, that our testing technique is valid. Besides, it is important to note that when identifying a foreign utterance, the American listener, as well as the French listener, used the auditory habits (and perhaps the articulatory habits) of their native language.

(b) It is noteworthy that our listeners were able to make the j/i distinction on the basis of duration alone, and did it clearly and regularly. We have no way of knowing whether they would make the distinction as clearly on the basis of other factors, such as voicing, rate of transition, overall intensity, but we can assume, from the significant results obtained, that duration is a very important correlate of the j/i distinction and offers a valid point of comparison between the two languages.

(c) Finally, the results of Fig. 7 show that, in the auditory perception of the j/i distinction, there is a marked divergence between the French subjects and the American subjects, which correlates well with the divergences observed earlier in the articulatory process.

The range of possible /j/ perception by French ears is about 0 cs to 10 cs in one case, 0 cs to 8 cs in the other. The range of /j/ perception by American ears is about 4 cs to 16 cs in one case and 4 cs to 14 cs in the other. French ears unanimously perceive a /j/ and hear Julia, Lilian only when the steady-state of formants lasts zero cs, that is, only when the second-formant transition dives back downwards as soon as the /l/ transition has finished rising. American ears perceive a /j/ and hear Julia, Lilian for durations of the steady-state formants of 0 cs, 2 cs, and 4 cs. French ears unanimously perceive an /i/ and hear Julie, Lily for durations of the steady-state formants of 10 cs. American ears do the same for durations of the steady-state formants of 14 cs or 16 cs.

The cross-over point from /j/ perception to /i/ perception is near 6 cs in the French judgements of French utterances and near 10 cs in the English judgements of English utterances, which means that, to perceive a /j/, American ears require the steady-state of formants to be 4 cs longer than is required by French ears. This difference of 4 cs seems to be minimal. It agrees with the limits of unanimous perception for /j/, which are 0 cs in the French judgements and 4 cs in the American judgements, but it is short when compared with the limits of unanimous perception for /i/, which are 12 cs in the French judgements and 18 cs in the American judgements -- a difference of 6 cs. The difference, then, might be better stated as ranging from 4 cs to 6 cs. Such a difference -- 4 to 6 cs -- is in closer agreement with the differences found, earlier in this study, on spectrograms and x-ray films. The spectrograms show a difference of 4 to 8 cs, and the films one of 1 to 2 frames, also meaning 4 to 8 cs.

Finally, another kind of comparison can be made by observing a single point of the result curves for both languages. For a duration of the steady-state formants of 8 cs, for instance, the French ears overwhelmingly perceive an /i/ and hear Julie, Lily, whereas, for the same duration, American ears overwhelmingly perceive a /j/ and hear Julia, Lilian. This last comparison is perhaps the most dramatic illustration of the divergence between the perceptual habits of American and French subjects in distinguishing /j/ from /i/.

SUMMARY

This is a cross-language investigation of the phoneme /j/ of Yes /jɛs/ or Hier /jɛr/. The objective is to specify the factors of phonetic interference which prevent a French speaker from correctly pronouncing an English word like Radio /redjo/ or an American speaker from correctly pronouncing a French word like Radio /Radjo/.

In English as in French, post-consonantal /j/ contrasts with other consonants (E., Cue /kju/, Clue /klu/, Crew /kru/, F., Quiet /kjɛ/, Clef /klɛ/, Craie /krɛ/), with zero (E., Cue /kju/, Coo /ku/, F., Quiet /kjɛ/, Quai /kɛ/), and with /i/ (E., It's Lilian her cousin /ɪts lɪljən hɚ kʌzən/, It's Lily an(d) her cousin /ɪts lɪli ən hɚ kʌzən/, F., Si Julia paraît /si ʒylja parɛ/, Si Julie apparaît /si ʒyli aparɛ/).

An attempt is made to relate the articulatory and auditory habits, for the pronunciation of post-consonantal /j/ in French and in English, to objective data obtained (a) by sound spectrography, (b) by cineradiography, and (c) particularly by the techniques of artificial speech synthesis which permit one to produce controlled changes in a single acoustic parameter at a time and to make linguistic judgements by ear of the effects produced by such changes.

(a) Spectrograms of comparable English and French sentences like The pianist from Vienna plays on the radio and Le pianiste viennois joue à la radio show five main acoustic differences. The English /j/ is more voiced, more intense, longer, lower in formant-two frequency, and slower in transition-two rate of change. In subjective terms, the English /j/ is more vocalic. (b) Motion picture x-rays of the English words Pianist, Vienna, Radio, and of the French words Pianiste, Viennois, Radio, show four articulatory differences. For the American /j/ the linguo-palatal constriction is wider, the articulatory motion is slower, it lasts longer, and it involves more consonantal anticipation and less vocalic anticipation. (c) Perceptual tests involving only variations in the duration of the /i/ formants in steady-state, show that French ears perceive a /j/ for shorter durations (ranging from 0 cs to 10 cs) than do American ears (durations ranging from 4 cs to 16 cs). When the steady-state formants have a duration of 8 cs, for instance, French ears overwhelmingly perceive an

/i/ and understand _Julie_, _Lily_, whereas American ears overwhelmingly perceive a /j/ and hear _Julia_, _Lilian_.

THE RADIOGRAPHY OF VOWELS AND ITS ACOUSTIC CORRELATIONS

How could one speak of vowels without recalling the phonetics' lesson in Molière's <u>Le Bourgeois Gentilhomme</u>:

PROFESSOR OF PHILOSOPHY: . . . There are five vowels or voices: A, E, I, O, U.
MONSIEUR JOURDAIN: I understand all that.
P. OF PHIL.: The vowel A is sounded by opening the mouth very wide, -- A.
M. J.: A, A. Yes.
P. OF PHIL.: The vowel E is sounded by bringing the lower jaw to the upper jaw, -- A, E.
M. J.: A, E; A, E. Bless me! How fine that is!
P. OF PHIL.: The vowel I is formed by bringing the jaws still closer together, and stretching the corners of the mouth toward the ears, -- A, E, I.
M. J.: A, E, I, I, I. That's true. Hurrah for science!
P. OF PHIL.: The vowel O is sounded by opening the jaws and drawing in the lips at the two corners, -- O.
M. J.: O, O. Nothing could be more true. A, E, I, O, I, O. It is admirable! I, O; I, O.
P. OF PHIL.: The mouth must be opened exactly like a round O.
M. J.: O, O, O. You are right. O, -- ah! what a fine thing it is to know something!
P. OF PHIL.: The vowel U is sounded by bringing the teeth together without entirely joining them, and protruding the lips outwardly, while bringing them narrowly together without actual contact: O, U.
M. J.: O, U, U; the truest thing that ever was, -- U.
P. OF PHIL.: Both your lips should be stretched out as if you were making a grimace; so that if you should ever want to make a face at any one and ridicule him you have only to say "U".
M. J.: U, U. True enough. Ah! why didn't I learn that in my youth?

In reading this celebrated passage one is surprised to find that the distance which separates popular notion from scientific truth is the same today as in the 17th century. For Monsieur Jourdain's Professor of Philosophy, "There are five vowels or voices: A, E, I, O, U," pronounced [a, e, i, o, y] as in French <u>sa</u>, <u>ses</u>, <u>si</u>, <u>sot</u>, <u>su</u>. In 1968, for the bourgeois from Philadelphia as well as the factory worker from Detroit, the number of vowels has not changed. Ask either one to recite the English vowels, and he will respond with: A, E, I, O, U, pronounced [éi, i, ái, óu, ju] as in English bay, bee, buy, bow, boo.

In French, as in English, there are in the majority of dialects not five but fifteen vowels; that is to say, fifteen classes of vocalic sounds (syllable nuclei) capable of effecting a change of meaning by simple substitution. The following sequences of minimal pairs, in which the vowel alone changes, will serve as an illustration. For French: lit, lut, loup, les, leu,[1] lot, l'air, l'heure, l'or, là, las, lin, l'un, lent, long. For English: keyed, kid, could, cooed, cade, curd, ked, cud, cawed, cad, cod; file, foul, foil.[2]

The Professor of Philosophy knows perfectly well that each vowel in a given language has a distinctive sound because the mouth assumes a different shape for each one. To make Mr. Jourdain understand that, the Professor of Philosophy limits his description to the outwardly visible organs -- to the widening of the jaw angle: "The vowel A is sounded by opening the mouth very wide . . . The vowel E is sounded by bringing the lower jaw to the upper jaw," and to the rounding of the lips: "The vowel U is sounded by bringing the teeth together without entirely joining them, and protruding the lips outwardly, while bringing them narrowly together without actual contact: U."

Mr. Jourdain was delighted by these summary notions of phonetics. How much greater his wonder would have been had the Professor of Philosophy placed him in front of an x-ray tube and shown him on a television screen what takes place <u>inside</u> his mouth, from the incisors to the pharynx and from the velum to the larynx, during the articulation of these same vowels. Thanks to the recent invention of light intensifiers it is possible today to photograph almost invisible radiographic images (reducing exposure to radiation to an infinitesimal degree) -- the intensifier increases the intensity of the images by a factor of 3000 and makes it visible to the camera. This is done in the same way as when the amplitude of acoustic waves is increased in a radio to make them audible to the ear.

In a well equipped laboratory of phonetic research, any <u>bourgeois gentilhomme</u>, whether curious or scientifically minded, can not only study on a television screen the articulatory gestures, but make a film of these movements and analyze them at leisure. Thanks to special projectors, he can see the film at normal speed, while hearing the speech sounds that were automatically recorded on the film, or see it in slow motion without losing the sound. He can even stop at each frame without time limitations and trace sketches of interesting images.

It was by tracing cineradiographic film images projected on opaque glass that the vowel profiles of Figure 1 were obtained. Note that these profiles do not involve posed photographs, but a selection of frames from cinematographic images made during the actual and natural pronunciation of words, and showing the most characteristic movement of vocalic opening. The speaker for these films is a Frenchman raised in the Loire valley, and without dialectal peculiarities. In this figure, therefore, we find vocal cavity shapes that are sufficiently representative of

Northern French. Now let us see what they can teach us.

We shall first glance briefly at Figures 1 and 2, then we will examine in turn the different ways of classifying the vowels from an articulatory, acoustic, and perceptual viewpoint. We shall see that one classification is purely practical, whereas the others have the advantage of explaining the relation between the physical aspect (acoustic) and the physiological aspect (articulatory).

Figure 1 presents the articulatory aspect, exclusively, and Figure 2 the acoustic aspect. The articulatory positions of Figure 1 are, therefore, those which produce the resonance notes of Figure 2, and the notes in turn are responsible for the perceptual distinction among vowels.

ACOUSTICS OF VOWELS

The horizontal bands of Figure 2 represent acoustically the two main resonance notes of the mouth, viewed as a single cavity limited by the lips at one end and the vocal cords at the other.[3] In the terminology of acoustic phonetics, however, one does not speak of notes but of formants; the color of a vowel is characterized by two formants, marked here at the right of the horizontal bands, F_1 (first formant or lowest formant) and F_2 (second formant from the bottom). The dotted line indicates zero frequency. The formants which resonate in the oral cavity have a frequency which is not limited to a single note, as the width of the horizontal bands attempts to indicate. But we show (at left) only the frequency of the center of each band. Using this center value it is traditional

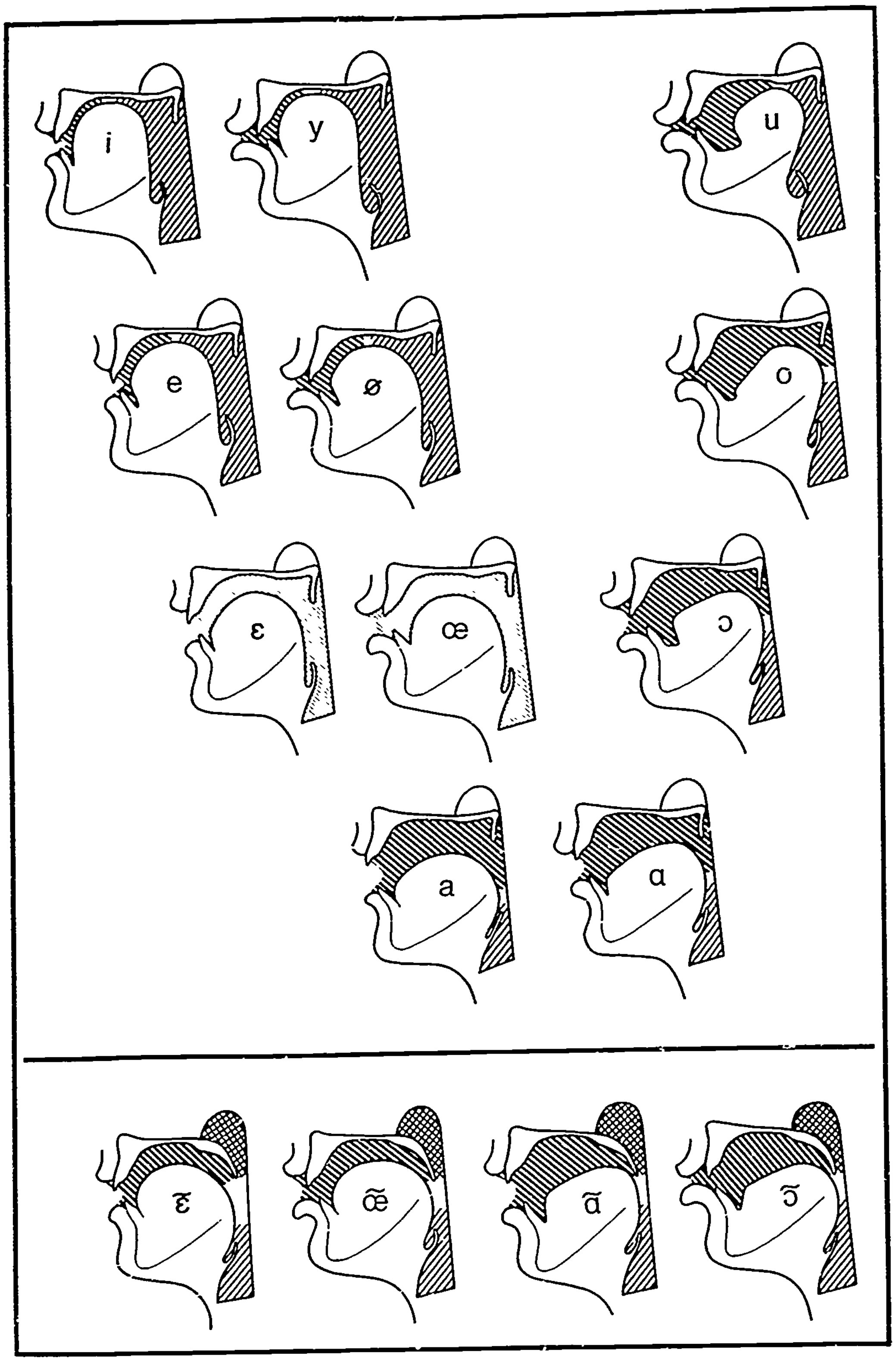


Figure 1. Lateral view of vocal tract strictures and cavities for French vowels.

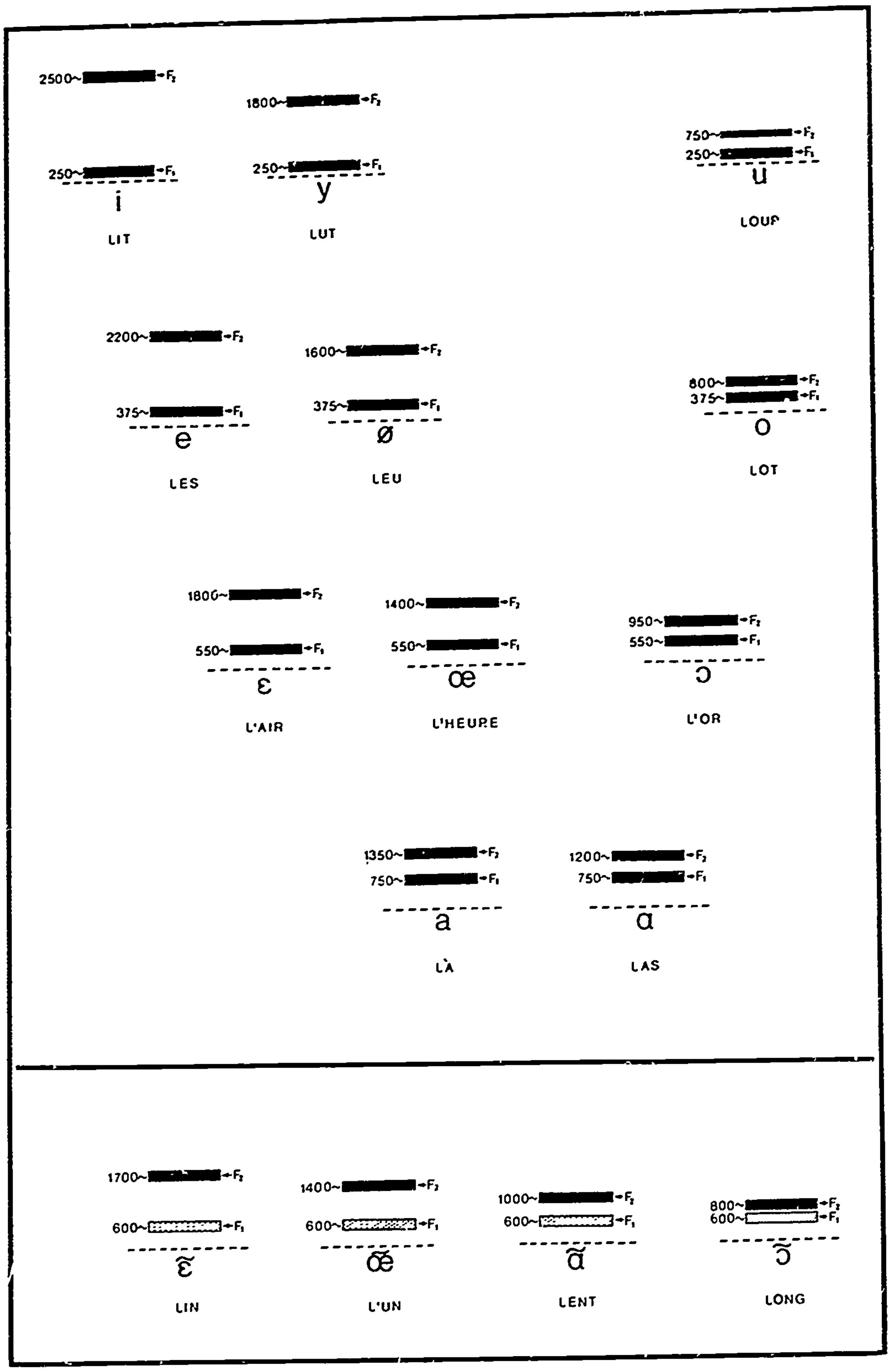


Figure 2. Distinctive formant frequency and intensity for French vowels.

to say, for instance, that the vowel /e/ is distinguished from the others by a musical chord (a formantic chord): F_1: 375 cps, F_2: 2200 cps; the vowel /a/ by a chord: F_1: 750 cps, F_2: 1300 cps, and so on.

Moreover, it is possible to hear the two formants separately without the help of any instrument. In tapping the neck, to the right or the left of the Adam's apple with the help of a pencil or a tap of the finger, the first formant is very well isolated from the second. Assuming succesively the articulatory positions of /a, ɔ, o, u/, having taken the precaution of closing the glottis with a slight laryngeal contraction, the tapping note descends quickly in four frequency-steps as do the first formants of these four vowels: 750 cps, 550 cps, 375 cps, 250 cps. The same series of descending notes can be heard when assuming the successive positions /a, ɛ, e, i/ or /a, œ, ø, y/. On the other hand, passing from /ɛ/ to /œ/ to /ɔ/, the tapping note remains more or less the same; correspondingly the first formant of these three vowels is the same.

The isolation of the second formant can be done by whispering the vowels in a quiet, well-isolated room. In passing from /i/ to /y/ by a very slow rounding of the lips, it is possible to hear the whispered note descending almost a fifth (from 2500 cps to 1800 cps). Then in passing from /y/ (1800 cps) to /u/ (750 cps) by slowly retracting the tongue towards the pharynx, this descent of the whispered note is extended beyond an octave, so that the note for /u/ (750 cps) is almost two octaves lower than the one for /i/ (2500 cps). The ascending or descending series can naturally be changed at will. Beginning with /a/, going toward

/i/ or /y/ the whispered note ascends, but going toward /u/ the whispered note descends, as does the frequency of the second formant.

VOWEL THEORY

In producing a vowel, man combines, therefore, two acoustic functions: a source and a resonance -- the source (at the vocal cords) produces a complex tone in a very wide frequency-band, and the resonance (in the cavities above the vocal cords) filters this comple.. tone allowing only the passage of narrow frequency-bands (formants) which coincide with the resonance notes of the resonator (or resonators) formed by the mouth from the vocal cords to the lips. In spoken-aloud speech, the source is represented by the vibrating vocal cords producing a rich series of harmonic tones (all pure, or sinusoidal), all simple multiples of the fundamental tone (melody or intonation of speech); the formants, then, are the result of only those harmonics which the filtering process of the oral cavities have let pass. In whispered speech, the source is represented by tightened and immobile vocal cords producing a white noise (non-periodic sound) in a very wide frequency-band when the air from the lungs is forced through the glottal slit; the formants, in that case, are narrow bands of <u>noise</u> which the filtering process of the oral cavities have let pass.

What counts in vowel perception is, therefore, not the harmonics (since there are none in whispered speech) but the formants, or frequency-bands, determined by the shape of the mouth. Whether the formants are composed of periodic sound as in the harmonics of the vibrated voice, or of non-periodic sound (turbulent), as in the noise of the whispered voice, they always fulfill their linguistic function in distinguishing between one vowel and another.

FORMANTS OF MEN, WOMEN, AND CHILDREN

Generally, formant frequencies are cited as absolute values. This is not altogether correct. The frequencies of Figure 2 are those of a man with average pitch, that is to say of approximately 120 cps. The formants of women are slightly higher -- on the average from 5 to 15 per cent higher according to the vowel; those of children are up to 25 per cent higher still. These differences are due to the dimensions of the oral cavities which are smaller for women than for men and still smaller for children. From the perceptual viewpoint the human ear is accustomed to identifying the vowels in relation to pitch. On the average, the female voice is higher than the male voice by about an octave (about 240 cps instead of 120 cps for men), because the feminine vocal cords vibrate approximately twice as fast. If a high voice pronounces, for example, an /a/, we understand /a/ only if the formants are slightly higher than those of a low voice for the same vowel. The following experiment on the speech synthesizer is based upon this perception theory. In synthesizing the word pomme [pɔm] with a fundamental according to a male voice, the two vowel formants are at 550 cps and 950 cps respectively, as for the /ɔ/ in Figure 2. To the degree that the fundamental is gradually raised without changing the formant frequencies, the meaning of the word tends to change from <u>pomme</u> with an open [ɔ] to <u>paume</u> [pom] with a closed [o]. The meaning of the word becomes ambiguous if the fundamental has been raised to a point between one and two octaves. In order to re-establish the vowel color that corresponds to <u>pomme</u>, the frequency of the two formants must be slightly raised. This experiment is favored by the fact that the two formants rise almost in parallel fashion in going from /ɔ/ to /o/.

X X X

After these fundamental notions about the acoustic properties of vowels, let us return to Figure 1 which presents their articulatory aspect.

We must first identify in our sketches the organs which come into play in the formation of a vowel. Let us proceed from left to right. The lips can be flat against the teeth or protruded, and they can be spread or rounded according to whether the corners of the lips are distant from each other or close together; they follow in other respects the widening of the jaw angle. In French, the widening implies the flattening of the lips; the rounding, the protruding of the lips.

The upper and lower incisors are located respectively in the upper maxillary and lower maxillary bone (mandible). We can see in the drawings, inside the chin, the tip of the lower maxillary. When the mouth opens, this maxillary pivots on its point of attachment to the skull (condyle of the mandible), while the upper maxillary remains immobile. This condyle is visible by x-ray, just to the right of the cushion of Passavant which forms the point of contact between the velum and the pharyngeal wall, when the velic passage is closed to prevent nasalization.

To the right of the upper lip, in the drawings of Figure 1, a bone which is attached to the upper maxillary and which takes up more than half of the complete palate, near the middle of the palatal ridge, we notice a thickening of the roof of the palate. It shows where the bone of the palate ends and the muscular membrane of the velum, also called the soft palate, begins. In the x-rays of the oral vowels of Figure 1, one has the impression that the velum ends in a right angle; in reality, the wide end of the velum lies against the vertical wall of the pharynx to permit closure of the velic passage so that the nasal and oral cavities can communicate; and the appendage which hangs vertically is the projection, called "uvula," which can be seen between the tonsils when opening the mouth widely to say Ah. This uvula vibrates and periodically touches the tongue during the pronunciation of the uvular R.

With the nasal vowels, seen at the bottom of Figure 1, this same velum lowers for the opening of the velic passage and permits the small rhino-pharyngeal cavity (indicated by the cross-ruled area) to combine its resonance with that of the oral cavity for the distinctive effect of nasality.[4]

The vertical line to the right is the wall of the pharynx. The pharynx is the rear part of the mouth, or, more simply, the throat. Its forward boundary is the root of the tongue which forms a vertical wall for /i/ and a pharyngeal constriction for /a/.

The horizontal line which defines the base of the pharynx represents, for four-fifths on the left side, the vocal cords (between which is the glottis). These cords, in the center of the larynx, are at the upper end of the trachea which conducts the air pressure from the lungs, pressure which is controlled by the intercostal muscles.

To the right of the vocal cords, extending from the pharyngeal wall, is the narrow entrance to the esophagus which takes food to the stomach. The little flap which rises against the base of the tongue is the epiglottis whose role it is to cover the glottis during the process of swallowing and to direct the food toward the esophagus.

Finally, the tongue, key to phonation because of the variety of forms it is able to assume, is generally divided into tip, blade (upper side adjacent to the tip), dorsum (front, back) and root. We mean by root that part which faces the pharyngeal wall, by dorsum that which faces the palate, and by tip that which faces the incisors or the alveoles; but it is possible, in retroflex sounds, to see the tip move toward the palate.

THE TONGUE-HUMP: INADEQUACY OF THE PHONETIC TRIANGLE

The first phoneticians, from Paul Passy, creator of the phonetic alphabet, to Daniel Jones, his successor as president of the International Phonetic Association, established the traditional phonetic triangle based upon the highest point of the tongue.

In order to envision this triangle, one has only to imagine that the articulatory profiles of /i/, /a/, and /u/, (Figure 1) are superimposed in such a way that the palates coincide. Then the highest point of the tongue, in relation to the palate, forms a triangle with /i/ at upper left, /u/ at upper right, and /a/ at bottom center. The other vowels can be inserted more or less arbitrarily between /a/ and /i/ or between /a/ and /u/. According to Figure 1, the vowels /e, ø, ɛ, œ/ would each have their highest point of the tongue close to the /a-i/ line, and the vowels /o-ɔ/ would have theirs close to the /a-u/ line.

This phonetic triangle has rendered great service in the field of historical phonetics. It explains with rare simplicity the tongue positions in two dimensions: front-back and high-low (or closed-open if one thinks in terms of the separation of the jaws which accompanies the lowering of the tongue). Unfortunately, the articulatory triangle has shortcomings with respect to the acoustics and the perception of vowels. First of all, it only takes into account the tongue; consequently it classifies /y/ at the same point as /i/ (these two vowels having nearly the same tongue position), /ø/ at the same point as /e/, etc. Furthermore, and this is more serious, it does not indicate what is most relevant to linguistic perception -- the highest point of the tongue has no direct relationship with the frequency of the acoustic resonances which distinguish vowels from one another perceptually. For the identification of vowel /a/, for instance, it is not the low level of the tongue-hump which counts acoustically, but rather the pharyngeal constriction which is formed between the root of the tongue and the wall of the pharynx. It is the place and narrowness of this constriction which is critical. This constriction separates the mouth into two cavities each of which favors a certain resonance (without being entirely independent as long as they communicate). It is, therefore, as we shall see later when considering the mouth as a cylinder, the distance which separates this constriction from the other constrictions at the lips and at the glottis which explains the frequency of the first and second formants.

The x-rays of Figure 1 fortunately allow us to make other classifications better related to the now known acoustic reality of vowels.

TONGUE CONSTRICTIONS

Starting with the profile, in Figure 1, and proceeding to the right, one can observe that the tongue constrictions circle the walls of the mouth -- the constriction of /a/ is low in the pharynx, that of /ɑ/ is a little higher, toward the middle of the pharynx, those of /ɔ/ and /o/ are higher still, in the upper portion of the pharynx; for /u/ the constriction reaches the velum, for /y/ and /i/ it advances up to the hard palate, for /ø/ and /e/ it widens and draws back slightly toward the pharynx, from where the /a/ constriction started.

As for /œ/ and /ɛ/, these vowels have, so to speak, no constriction. For them, the mouth assumes the shape of a cylindrical tube of somewhat uniform diameter. The tube for /ɛ/ is the shorter and more open of the two. It is for /œ/, the neutral vowel, that the mouth best resembles the simple shape of a uniform tube. Figuratively speaking, one can consider the vowels /œ/ and /ɛ/ as a bridge between those which have a constriction at the palate and those which have one at the pharynx, thus closing the circle of vocalic constrictions.

The nasal vowels (lower row in Figure 1) have this in particular that they always form their tongue constriction at the pharynx, just below the tip of the lowered velum. We shall see later the significance of this fact.

LIP CONSTRICTION

The constrictive play of the lips, although less varied than that of the tongue, is also very important. As one can see in the profiles of Figure 1, it is principally the labial constriction (rounding) which changes /i/ to /y/, /e/ to /ø/, /ɛ/ to /œ/, /ɛ̃/ to /œ̃/, and /ɑ̃/ to /ɔ̃/ and, at least partially, modifies /a/ to /ɑ/. This labial constriction is, therefore, comparable in importance to the displacement of the tongue constriction along the walls of the pharynx and palate which plays the principal role in changing /y/ to /u/, /ø/ to /o/, /œ/ to /ɔ/, and /a/ to /ɑ/.

VELIC CONSTRICTION

The velum when lowered forms a constriction called velic which contributes greatly to the change from /ɛ/ to /ɛ̃/, /œ/ to /œ̃/, /ɔ/ to /ɔ̃/, and /ɑ/ to /ɑ̃/. We intentionally say "contributes" because the articulatory position of the nasals is quite different from those of the corresponding orals, as is well shown in the x-rays of Figure 1. Thus, the front cavity for /ɔ̃/ is more like the one for /o/ than like the one for /ɔ/; the back cavity for /ɑ̃/ bears more resemblance to the one for

/ɔ/ than to the one for /ɑ/. Moreover, duration plays a certain role in the nasal/oral distinction.

GLOTTAL CONSTRICTION

Finally, the vocal cords themselves form a constriction, whether in the position for vibration or for whispering. One becomes aware of this when tapping the throat close to the Adam's apple in order to hear the resonance note of the first formant for /ε/, for example. During this procedure, if one opens the glottis as if for breathing, the resonance note disappears, or becomes so low that one can no longer discern it. Such deterioration of the resonance process occurs because the pharyngeal cavity has now been lengthened by the trachea and the volume of the pharyngeal cavity is immeasurably increased.

Furthermore, the glottal constriction plays a role when falling or rising with the entire larynx -- lowering the larynx increases the total length of the mouth cavity or of the pharyngeal cavity. One sees in Figure 1, for instance, that the glottis is considerably lower for /u/ than for /a/.

RESONANCE CAVITIES AND FORMANTS

The tongue constrictions which we have just described tend to divide the mouth into two resonance cavities. Theoretically, since these cavities communicate, they do not resonate separately, and each modification of one of the two affects the frequency of both. But practically speaking, one may consider the frequency of the first formant as related to the back cavity (pharynx) and that of the second formant to the front cavity. The more narrow the constriction the more negligible is the element of error. (In the x-rays of Figure 1, we distinguished between front and back cavities by using different hachures.)

The resonance formula which applies here is simple: <u>the larger the cavity and the smaller and longer its opening, the lower its note of resonance</u>; <u>and conversely</u>. Let us add immediately that due to the play of compensations, a cavity of little volume and small opening (two opposing effects) can have the same frequency as a cavity of large volume and large opening (also two opposing effects).

CORRELATIONS BETWEEN THE BACK CAVITY AND THE FIRST FORMANT

According to Figure 1, the back cavities for /i/ and /y/ (top row) have almost the same large volume and small opening (tongue constriction). Accordingly, in Figure 2 the first formants of these two vowels have about the same frequency.

The same observation applies to the vowels /e/ and /ø/: their back cavities are similar as are their first formants.

It also applies to the nasal vowels /ɛ̃, œ̃, ɑ̃, ɔ̃/ (bottom row); their back cavities are similar (Figure 1) as are their first formants (Figure 2).

In the series /u, o, ɔ, ɑ/ the volume of the back cavity decreases regularly. The frequency of the first formant, however, increases in this same order, thus illustrating the fact that the smaller the cavity, the higher its resonance note. The series /y, ø, a/ and /i, e, a/ illustrate the same law. Note here that the /a/ type vowels have the smallest back cavity and of all the vowels the highest frequency for the first formant. (The fact that the back cavity of /u/ and /o/ is smaller than that of /y/ and /ø/, respectively, although their first formants hardly differ, is attributable to the lowering effect of the front cavity which is large and closed, hence has a low resonance frequency.)

CORRELATIONS BETWEEN THE FRONT CAVITY AND THE SECOND FORMANT

In examining Figure 1, we will first make comparisons among certain profiles. In passing from /i/ to /y/, the rounding of the lips considerably reduces the opening of the front cavity and slightly increases its volume. Theoretically, both effects should contribute to lowering the resonance note of this cavity. Figure 2 shows, in fact, that in passing from /i/ to /y/, the frequency of the second formant is considerably lowered.

In passing from /y/ to /u/ the retraction of the tongue constriction increases the volume of the front cavity which should lower the resonance note of this cavity even further. Figure 2 shows, in fact, that in passing from /y/ to /u/ the frequency of the second formant is greatly lowered.

The same logic can be applied to the second row of profiles, but with less pronounced effects. Let us compare /e/ to /ø/. The rounding of the lips for /ø/ lessens the opening and enlarges the volume of the front cavity. This should theoretically lower its resonance note. Accordingly, the second formant is lower for /ø/ than for /e/.

Let us compare /ø/ to /o/. The retraction of the tongue constriction for /o/ increases the volume of the front cavity. This should result in a lower resonance note for /o/ than for /ø/. Thus, as is shown in Figure 2, the second formant is lower for /o/ than for /ø/.

The influence of the lips is again visible in the nasals. In comparing /ɛ̃/ to /œ̃/, we see that lip rounding reduces the opening and increases the volume of the front cavity. Accordingly, a lowering of the second formant frequency appears in Figure 2.

Let us compare /ɑ̃/ and /ɔ̃/. Lip rounding strongly reduces the opening of the front cavity. This correlates in Figure 2 with a lowering of the second formant frequency (in spite of a slight decrease in front-cavity volume).

Let us now compare the front cavity of the vowels which are at the three corners of the triangle.

The front cavity of /i/ is very small and since the corners of the lips are spread, the opening of the cavity is medium. It is natural, therefore, that the second formant of /i/ be higher than for any other vowel.

On the contrary, the front cavity for /u/ is large and its opening very small. One should, therefore, expect the second formant of /u/ to be very low; it is, in fact, the lowest of all, but followed closely by /o/ (which is distinguished from /u/ by the first formant more than by the second).

What can now be said about /a/? It offers a case of compensation. Whereas for /i/ there is a concordant effect between the smallness of the front cavity and the largeness of the opening, and for /u/ a concordant effect between the largeness of the front cavity and the smallness of the opening, for /a/ there is an opposing effect between the largeness of the front cavity and the largeness of the opening. The fact that the front cavity for /a/ is very large (which should result in a very low note) is largely compensated for by its immense opening (which causes the frequency to rise). Thus, we have an explanation for the second formant of /a/ being intermediate in relation to /i/ and /u/.

THE CHORD OF THE TWO RESONANCE NOTES

A brief look at the two cavities of the vowels in Figure 1 should now enable us to predict the formant frequencies of those vowels.

The vowel /i/, formed by a very large and closed back cavity and a very small and open front cavity, should be characterized by the combination of a very low and a very high note. The vowel /u/, formed by two cavities which are both large and closed, should be characterized by two low notes. The vowel /a/, formed by a small and closed back cavity (two opposing factors), and a large and open front cavity (again two opposing factors) should be characterized by two medium notes. These assumptions are confirmed in Figure 2.

The resonance notes of the other oral vowels can be explained by the intermediate positions they have in relation to the above three.

All the nasal vowels have a medium and slightly closed back cavity with a resonance note that is comparatively high (according to the scale of the first formants). They are distinguished from one another by the front cavity which is medium and open for /ɛ̃/, medium and closed (corners drawn together) for /œ̃/, large and open for /ɑ̃/, large and closed for /ɔ̃/. This should produce four different resonance notes ranging from middle-high for /ɛ̃/ to low for /ɔ̃/. Figure 2 confirms that assumption -- the nasal vowels are distinguished from each other by their first formant alone.

The nasal vowels, by lowering the velum, add a third cavity to the resonance system. This nasal cavity acts upon the back cavity (pharyngeal): it damps the resonance, and even cancels some of its harmonics by means of counter-resonance. The perception of nasality is, thus, simply the result of an unbalance in the relative intensities of the formants -- the first formant is now much weaker than the second. (To show this lowering of the first-formant intensity these formants are shaded in grey, in Figure 2, in contrast to the solid black of their oval counterparts.)

THE MOUTH SEEN AS A TUBE

In the two sections on the correlation between mouth cavities (front, back) and formants (second, first), we intentionally did not examine the particular aspect of the vowels /ɛ/ and /œ/ for which the mouth is not clearly divided into two cavities. In order to explain the frequency of the formants of these two vowels, we have to call upon a totally different theory which, though more exactly applicable than the others, is too demanding from the viewpoint of mathematical knowledge to be more than summarily presented here.

According to this theory, all vowels are explained as modifications of a neutral vowel which is produced in a cylindrical tube of uniform diameter. (We can see in Figure 1 that for /œ/ and /ɛ/ the shape of the mouth resembles that of a tube.) In a long, uniform tube, the resonance frequency depends solely upon the length of the tube, not upon its width or its volume. From a tube of 17.5 centimeters (average length of a man's mouth) closed at one end (the glottis) and open at the other (the lips), result resonances of 500 cps (first formant), 1500 cps (second formant), 2500 cps (third formant), 3500 cps (fourth formant), etc., which correspond to one-quarter of a wave length, three-quarters of a wave length, five-quarters of a wave length, etc., of a 17.5 cm tube in which the speed of sound is 3500 cm per second. The frequencies of 500 cps and 1500 cps are close to those of the first two formants of /œ/: 550 cps and 1400 cps or of /ɛ/: 550 cps and 1800 cps.[5]

Here you have to recall your first physics course of long ago. The resonance of one-quarter of a wave length (first formant) forms a node at the closed extremity (glottis) of the tube and a loop (anti-node) at its open extremity (lips). The resonance of three-quarters of a wave length (second formant) forms two nodes and two loops, one node at the closed extremity (glottis), one loop at one-third of the tube length, one node at two-thirds of the tube length, and one loop at the open extremity (lips). The formant frequency of every vowel can be found by applying the following laws separately to its two lowest modes of resonance (one-quarter of a wave length and three-quarters of a wave length): <u>The frequency of a mode of resonance rises</u>

or falls according to whether the constriction approaches a node or a loop, respectively. Thus, the pharyngeal constriction of /a/ (see Figure 1) is near a node (near the glottis) of the *first* mode of resonance (one-quarter of a wave length); hence, the first formant is higher than for the neutral vowel, that is to say, higher than 500 cps. This pharyngeal constriction is at the same time near a loop of the *second* mode of resonance (three-quarters of a wave length); hence, the second formant is lower than the one of the neutral vowel, that is to say, lower than 500 cps. Accordingly, the formant frequencies of /a/ are approximately 750 cps and 1300 cps (Figure 2).

For /e/ -- another example -- the palatal constriction is near a loop of the first mode (near the lips); hence, the first formant is lower than 500 cps. At the same time, the /e/ constriction is near a node of the second mode of resonance; hence, the second formant is higher than 1500 cps. Accordingly, the formants of /e/ are approximately at 375 cps and 2200 cps, in Figure 2.

DISTINCTIVE ACOUSTIC FEATURES

According to Figure 2, a vowel can be distinguished from another by one, two, or three acoustic features.

A single distinctive feature. Example: /e/ is distinguished from /ø/ by the frequency of the second formant alone. This condition exists in all the rows of Figure 1 -- the vowels of the same row are distinguished from one another solely by the frequency of the second formant. Thus, acoustically speaking, /i/ is distinguished from /u/ by a single feature, /ɛ̃/ from /ɔ̃/

by a single feature: the second-formant frequency. This creates a problem since, from the traditional viewpoint, such vowels are distinguished from one another by at least two articulatory features: the rounding of the lips and the retraction of the tongue! In perceptual terms, a question arises which has not been answered: Does the brain perceive the color distinction between two vowels <u>directly</u> by means of the formant-frequency waves which strike the ear drums, or <u>indirectly</u> by reference to the articulatory features which produce those formant frequencies?[6]

The distinction 'nasal vowel/oral vowel' can also be considered as principally determined by a single acoustic feature, the intensity of the first formant, as long as the nasal vowel approximates the formant frequencies as its oral counterpart. This is to a certain extent the case for the pairs /e/ɛ̃/ and /œ/œ̃/ but much less for the pairs /ɔ/ɔ̃/ and /a/ɑ̃/.

<u>Two distinctive features</u>: From one row to another row, in Figure 1, the oral vowels are generally distinguished by two acoustic differences. Thus, the formants composing /y/ are at 250 cps and 1800 cps, those for /ɔ/ at 550 cps and 950 cps. In two particular cases of Figure 2, however, the distinction seems mainly attributable to the first formant alone. These are the pairs /y/ɛ/ and /u/o/.

<u>Three distinctive features</u>: Except in the cases mentioned above where a nasal vowel has an oral counterpart, all nasal vowels are distinguished from oral vowels by three acoustical differences: the frequency of the first formant, the frequency of the second formant and the intensity of the first formant.

Let us mention a fourth feature (but merely in passing, since in French it is either unstable or secondary). Vowel duration distinguishes maître from mettre; besides, it contributes to the distinction of paume from pomme, tâche from tache, of cinq from sec, but then, duration is conditioned by vowel color, hence, not separately distinctive.[7]

It goes without saying that the auditory confusion between two vowels is reduced in direct proportion to the number of features which separate them. Thus, in a crowded room where everyone is speaking at the same time, bulle is more easily confused with boule than with balle.

Thus, in French, four acoustic features serve to distinguish fifteen vowels from one another: the frequency of the first formant, the frequency of the second formant, the intensity of the first formant and the duration of the two formants. The acoustic facts are very clear; the articulatory facts and their role in perception are unfortunately less clear.

X X X

Let us summarize. Using x-rays of vowels taken from cineradiographic films of speech which show the entire oral cavity from the vocal cords to the lips, we have confronted the traditional vowel classifications with new descriptions justified by these x-rays. It has enabled us to show that the vocalic formants which are responsible for the perception of vowels have no direct relationship to the highest point of the tongue (traditional classification), but are explained (a) by the place of constriction in the mouth and (b) by the shape and volume of the two main mouth-cavities (front and back) or (c) by envisaging the mouth as a uniform tube and the vocalic positions as modifications of the tube through the formation of constrictions.

Thus, the examination of cineradiographic frames serves to explain the real relationship that exists between the acoustic aspect and the articulatory aspect of vowels, and allows one to judge objectively the superficial notions passed on by tradition.

FOOTNOTES

[1]Medieval term for wolf, retained in the expression *à la queue leu leu*.

[2]The Professor of Philosophy's error, like that of Rimbaud in the sonnet *Vowels*, stems from the confusion between 'letters' and 'vowels'. In writing the 15 different vocalic sounds of French, one makes use of only 5 characters -- the Latin letters a, e, i, o, u -- which explains the chaos of our spelling and the difficulty which our poor children have in learning it.

[3]In *French Review* for May, 1948 (Vol. 21, pp. 477-484) appear spectrograms of oral vowels of a French speaker, made on the first spectrograph from the Bell Telephone Laboratories, in the spring of 1947. They mark an historic date; the very first publication of spectrograms with a linguistic intent, and they have been reproduced in *PMLA* for September, 1951 (Vol. 66, pp. 864-875) and in *Studies in French and Comparative Phonetics*, The Hague, Mouton, 1966, p. 239.

[4]For an acoustical, articulatory and percptual discussion of nasality, see French Review, October, 1965 (Vol. 40, pp.218-223).

[5]We have not mentioned the third formant because it varies only slightly -- it remains near 2500 cps for all the vowels, except for /i/ where it reached 3000 cps. The third formant, thus, is not distinctive, or at least only slightly. The higher formants have really no linguistic function -- rather they determine the quality of the voice.

[6]For a discussion of this question, see: "Acoustic or articulatory invariance?" Glossa (Vol. 1:1, 1967, pp.3-25).

[7]The problems of vocalic duration are discussed in French Review, October, 1959 (Vol. 32, pp. 547-553), in Studies in French and Comparative Phonetics, The Hague, Mouton, 1966, pp. 105-141, and in Comparing the Phonetic Features of English, German, Spanish, and French, Philadelphia, Chilton Books, 1965, pp. 63-66.

PHARYNGEAL FEATURES IN THE CONSONANTS OF GERMAN, SPANISH, FRENCH, AND AMERICAN ENGLISH

This investigation involves eleven cases of consonant pharyngealization from five languages, three in German, one in Spanish, one in French, one in American English, and five in Arabic. The noted pharyngeals of Arabic are used as a reference to evaluate the extent of pharyngealization in the other four languages.

INTRODUCTION

A pharyngeal articulation is one in which the root of the tongue assumes the shape of a bulge and is drawn back toward the vertical back wall of the pharynx to form a stricture. This radical bulge generally divides the vocal tract into two cavities, one below extending from the stricture to the glottis, the other above extending from the stricture to the lips.

The best example of a pharyngeal articulation is the vowel /a/. As can be seen on Fig. 1, Row 1, Frame 2, for an /a/ the tongue root bulges toward the back wall of the pharynx, separating the vocal tract into a small cavity below the approximation and a very large cavity above it. The pharyngeal cavity below the bulge being small, its note of resonance is high (the larger and the more open a cavity, the higher its resonance frequency) -- in fact, the first formant of /a/ has the highest first-formant frequency of any vowel, an acoustic feature which reflects the fact that the pharyngeal cavity of /a/ is the smallest back cavity of any vowel. [This is a practical simplification which could suffice here. Actual acoustic theory would say that for /a/ the first mode of resonance -- or first "formant" -- is the highest of all vocalic first modes because its stricture is the nearest one to the closed end of the vocal tract tube (the glottis) and the farthest one from the open end of the tube (the lips). Acoustically, the mouth may be seen as a uniform tube, open at one end and closed at the other. For such a tube the first mode of resonance is that of the one-quarter wave length whose node is at the closed end and whose anti-node is at the open end. When a stricture occurs in the tube, the resonance frequency of the whole tube is raised or lowered depending upon whether the constriction is near the node which is at the glottis or near the anti-node which is at the lips. This may sound a little complicated, but it is basic if we are to understand what characterizes the acoustics of pharyngeal consonants.]

Since vowels are characterized by two formants, the second formant of /a/ must also be explained. The front (or mouth) cavity of /a/ being very large, the second formant would be very low were it not for the wide opening of the cavity, a factor that compensates for the large volume. The second formant of /a/, therefore, has a mid-low frequency. [Acoustic theory would put it another way and say that the second mode of resonance of the tube, for /a/, is near a loop, the loop of the three-quarter wave length that is closer to the glottis than to the lips.]

The pharyngeal vowel /a/, then, is characterized by a very high first-formant at about 750 cps and a rather low second-formant at about 1300 cps. First and second formants are, therefore, close together. This is the typical effect of pharyngealization, not only in vowels but in consonants as well. On spectrogram, consonants with a pharyngeal stricture can generally be recognized by a postvocalic rise of the first-formant transition and a postvocalic fall of the second-formant transition which bring the two formants close together; and in prevocalic position, naturally, the reverse is true -- the first-formant transition falls toward the first-formant frequency of the following vowel, and the second-formant transition rises toward the second-formant frequency of the following vowel. This is in accord with the recognized theory that vowels are perceived by static conditions -- steady-state frequencies of their first and second formants; whereas consonants are perceived by dynamic conditions -- rapid shifts of their formant frequencies, known as formant transitions.

Pharyngeal consonants are considered to be unusual speech sounds. Not all languages use them and those that do have very few of them. In theory consonant strictures can be produced at any place along the pharyngeal and palatal walls, from the vocal cords to the lips; in practice, however, the great majority of strictures are at the lips and the palate, not at the pharynx.

Furthermore, the remote place of the pharynx makes its articulations difficult to investigate. The pharynx can be observed only by x-ray, and until recently the image intensifiers, which determine the size of the articulatory area to be safely photographed, had too small a diameter to permit seeing more than a portion of the vocal tract, from profile. For this reason, if the lips showed, the pharynx did not, if the tongue tip showed, the tongue root did not.

Now, at last, larger image-intensifiers have been developed which make it possible to include the whole vocal tract in the frames of an x-ray film. Such equipment was used in this investigation to observe the motions of the root of the tongue -- the portion of the tongue which faces the pharyngeal wall, just above the vocal cords.

Native speakers of German, Spanish, French, English, and Arabic were used to pronounce, in front of our x-ray installation, sentences and minimal pairs illustrating the pharyngeal sounds of their respective languages. Cineradiographic films, with optical sound automatically recorded in the margin, were made, and they were studied frame by frame with the help of special projectors capable of projecting at normal speed or in slow motion.

Selected frames from those films are presented in Figs. 1 to 11. Tracings of each frame are made by transparency on enlargers. They show a profile view of the vocal-tract shape as it follows the outer curve of the tongue from the front teeth on the left to the glottis on the lower right. The horizontal line that limits the vocal tract at the lower-right end represents the vocal fold which generally appears horizontally as an elongated oval on the x-ray frames taken from profile. All films were taken at 24 frames per second; the time interval between each frame is therefore about 4 centiseconds (cs).

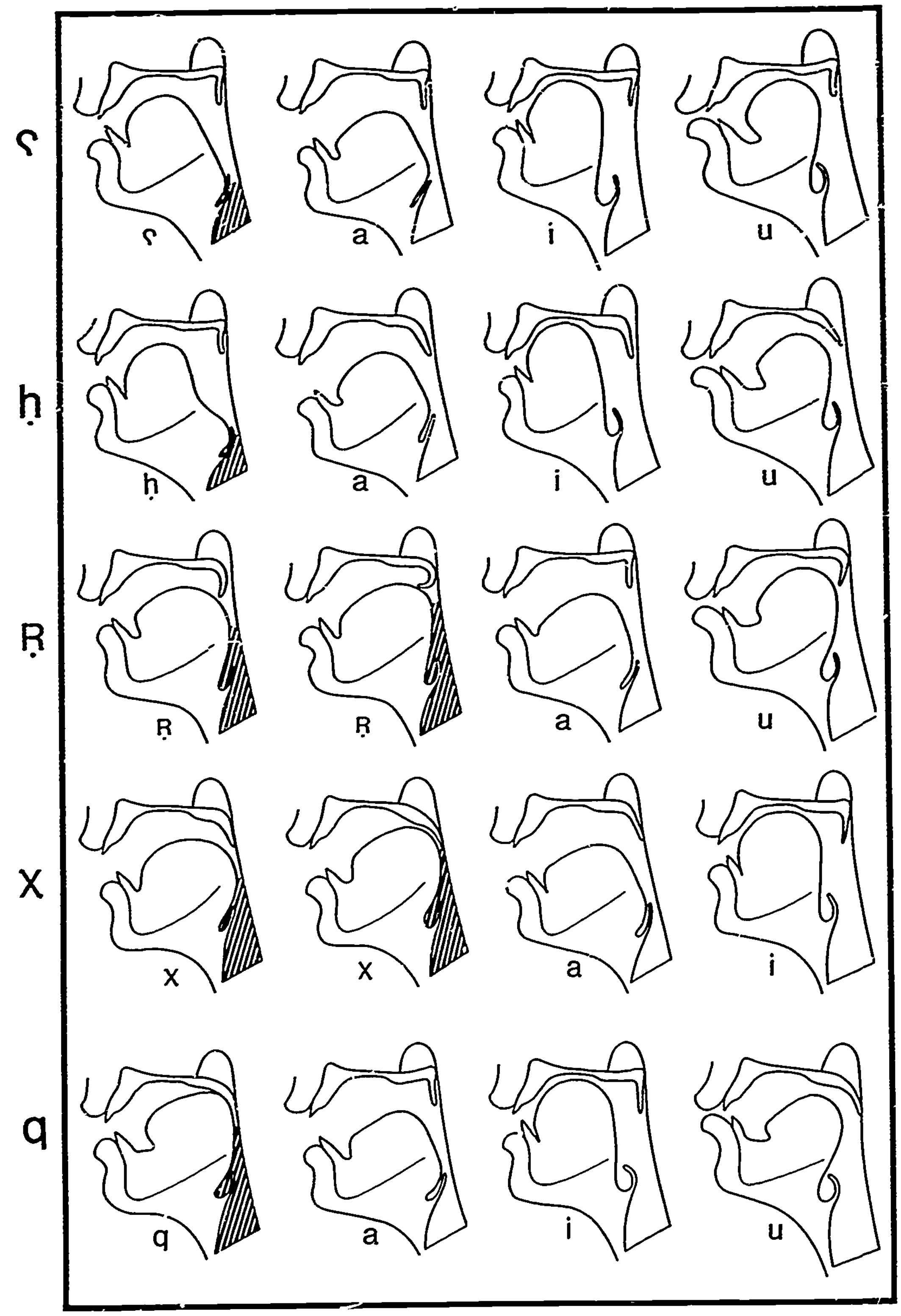
ARABIC

Let us begin with a brief description of the five pharyngeal consonants of Arabic which are presented in Fig. 1. A special article will be devoted to these Arabic sounds at the proper time -- here, we include them solely as a point of reference for the description of pharyngeals in other languages.

The symbols used here to represent the five Arabic pharyngeals are: /ɕ/, /ḥ/, /Ṛ/, /χ/, and /q/. The first and third are voiced, the others are voiceless; the first two are constricted in the lower pharynx, the last three in the upper pharynx; the first four are constrictives, the fifth one is a plosive. These five sounds are clearly distinctive and are not more than normally modified by the adjacent vowels. In the Arabic system of emphatic vs. non-emphatic consonants, in which the emphatics are more backed than their non-emphatic counterparts, /ɕ/ is generally considered the emphatic counterpart of /Ṛ/, /ḥ/ the emphatic counterpart of /χ/, and /q/ the emphatic counterpart of /k/ which is a palato-velar in Arabic, comparable to German or Spanish /k/.

The speaker who pronounced before the x-ray camera the Arabic sounds whose articulatory shapes are sketched in Fig. 1 is an educated native of Lebanon. The five consonants traced here were all in initial position, followed by /i/, /a/, or /u/. Each one of the five rows illustrates a different consonant. On the left in each row, the moment of maximal stricture in the constricting tongue-motion is shown. On the right in each

ARABIC



FIVE PHARYNGEAL CONSONANTS

Figure 1

row are two or three articulatory sketches of the vowels /a/, /i/, and /u/ as pronounced post-consonantally by the same speaker. These vowels are there to facilitate the immediate comparison of the pharyngeal cavities for the consonant and for the vowel and to permit one to visualize the motion of the tongue from the vowel to the consonant and back. The pharyngeal cavity of each consonant at the instant of maximal stricture is made visible by shading. When describing each consonant, we shall consider both the place of the stricture along the pharyngeal wall, and the volume of the pharyngeal cavity below that stricture. Generally these two features are correlated -- the lower the place of constriction, the smaller the pharyngeal cavity and the higher the first-formant frequency. The place of the stricture along the pharyngeal wall is also correlated with the volume of the _mouth_ cavity, above the stricture -- the lower the pharyngeal stricture, the larger the mouth cavity and the lower the second-formant frequency. But it must be added that the resonance note of these two cavities also depends upon the degree of closure at the jaws and the lips. Lip rounding, as might occur before /u/, for instance, lowers the resonance note of both cavities, but especially that of the mouth cavity, so that both formants have lower frequencies.

Row 1, left, shows that the voiced constrictive /ʕ/ is produced with a _very low_ stricture between the tongue root and the pharyngeal wall. This stricture is so low that it is even lower than for the vowel /a/ (next, to the right), which has the lowest stricture of all vowels. Furthermore, the front portion of the tongue shows that the tongue dorsum for /ʕ/ is higher and more fronted than for /a/, and that the jaws are closer. Now if we look at the two cavities of the vocal tract, we find that the pharyngeal cavity (below the tongue-root stricture) is even smaller for /ḁ/ than for /a/ (which has the smallest pharyngeal cavity of all vowels). A comparison with /i/ and /u/, to the right, shows how small the pharyngeal cavity of /a/ already is, as compared with those of /i/ and /u/ -- for /i/, the pharyngeal cavity is wide and long, it extends from the glottis to the palate; for /u/, it is not quite so wide as for /i/, but it is longer, the larynx being lowered. Naturally, with such a large volume of the pharyngeal cavity, the vowels /i/ and /u/ have very low first formants -- on the scale of first-formant frequencies, /a/ and /u, i/ are at the two opposite ends, /a/ at the highest frequency extremity, /u, i/ at the lowest. As to the mouth cavity for /ḁ/ (above the pharyngeal stricture and up to the lips), it is not quite so large as for /a/, but it is less open (a compensating factor).

In acoustical terms these cavity volumes and shapes indicate that the first formant should be higher for /ḁ/ than for /a/, and the second formant should be approximately the same for /ḁ/ as for /a/. At least this is the case when /ḁ/ is adjacent to /a/. When it is adjacent to /u/, coarticulation should reduce the high level of both formants to lower; when it is adjacent to /i/, coarticulation should reduce the high level of the first formant and raise the lower level of the second formant. All this is confirmed by the analysis of formants in spectrograms and by their synthesis on artificial-speech machines.

Row 2, left, shows that the voiceless constrictive /ḥ/ may well be regarded as the voiceless counterpart of the voiced /ç/, as generally assumed. Its pharyngeal stricture is *very low* (even lower than for /ç/), and its pharyngeal cavity is *very small* (even smaller than for /ç/). The pharyngeal stricture is also narrower for /ḥ/ than for /ç/, which is to be expected since, in the absence of voicing, the friction noise must be loud enough to carry the load of perception alone. The back of the tongue is cambered, as if the radical bulge toward the lower pharynx forced a compensating hollow above it.

These small differences between /ç/ and /ḥ/ are confirmed by the analysis of spectrograms. The first-formant transition of /ḥ/ is turbulent (non-periodic, inharmonic, noisy), and it rises slightly higher than the first formant of /ç/, which means considerably higher than the first formant of /a/.

Differences in the *mouth* cavity volume and in the consequent second-formant frequency of /ç/ and /ḥ/ do not seem to be significant.

A certain analogy with the American /r/ should perhaps be mentioned here. The peculiar tongue shape of /ḥ/, with its two bulges, one at the tongue dorsum toward the palatal ceiling, and another at the tongue root toward the pharyngeal wall, is remindful of the two constrictions which characterize the American /r/, as can be seen in Figs. 9 and 10. It has been shown elsewhere ("A dialect study of American *r*'s by x-ray motion picture," to appear in *Linguistics*) that the palatal stricture of the American /r/, when combined with a pharyngeal stricture, causes the *third* formant to lower extensively. Having noted the dorsal bulge of /ḥ/ and /ç/, we looked for a corresponding fall of the third formant on spectrograms. It is clearly present. The third-formant transition lowers regularly for /ḥ/ as well as for /ɋ/, but much less than for American /r/ -- which is comprehensible since the dorsal bulge is much less high for the two Arabic consonants than for the American one.

If we mention the third-formant effect and the dorsal bulge at this point, it is because later, when Figs. 9 and 10 are examined, the emphasis will be on the <u>first</u>-formant effect and the <u>radical</u> bulge.

Row 3, left, shows a typical shape of the vocal tract for the Arabic voiced constrictice /Ṛ/. This sound is represented by <u>two</u> sketches because it always includes two movements: first the tongue withdraws horizontally toward the pharyngeal wall to form a stricture near the middle of it, then the stricture rises along the pharyngeal wall, as if to permit the uvula to contribute a few trills which, in the case of our Lebanese subject, are not really interruptions of the air-stream but intermittent reductions in the intensity of the air-stream of a subdued nature. The uvular contribution is hardly noticeable on spectrograms, and does not even always take place. Its role seems to be limited to increasing the audibility of the sound. For this reason, it must be considered secondary. The primary factor in the perception of this /Ṛ/ is the <u>high</u> pharyngeal constriction and the volumes of the pharyngeal and mouth cavities. Like the /ɋ/ stricture, the /Ṛ/ stricture is wide for that of a consonant, but it is much higher on the pharyngeal wall than the /ɋ/ stricture, a factor which causes the pharyngeal cavity to be larger and the mouth cavity to be smaller than for /ç/. As a result the first-formant frequency is lower for /Ṛ/ than for /ç/ (but still high when compared with any other consonant), and its second-formant frequency is slightly higher (but still quite low). Compared with /a/, the /Ṛ/ stricture is higher in the pharynx; consequently its pharyngeal cavity is larger. This places the /a/ features between those of /ç/ and those of /Ṛ/, in Arabic -- the stricture of /a/ is higher than that of /ç/ and lower than that of /Ṛ/, along the pharyngeal wall; and the pharyngeal cavity of /a/ is larger than that of /ç/ and smaller than that of /Ṛ/. Similarly, the first-formant frequency for /a/ is lower than for /ç/ and higher than for /Ṛ/. Average frequencies of the first formant, for a male voice, must be in the vicinity of 750 cps for /a/, 1000 cps for /ç/, and 550 cps for /Ṛ/. These frequencies are all above 500 cps which is precisely the frequency of the first mode of resonance (first formant) of a uniform pipe of about 17.5 centimeters long, wide open at one end (the lips) and closed at the other (the glottis). According to acoustic theory, when a stricture is produced in such a uniform pipe, the frequency of the first mode of resonance of the whole pipe rises above 500 cps if the stricture is nearer to the closed end (the glottis) than to the open one (the lips); and conversely it falls below 500 cps if the stricture is nearer to the open end than to the closed end. In other words, according to acoustic theory, the front of the mouth tube being open, as for /a/, and no other stricture being introduced, the first-formant frequencies of /Ṛ/ (550 cps), /a/ (750 cps), and /ḉ/ (1000 cps) are above 500 cps because they are located in the back half (the pharynx) of the vocal tract; and the closer those strictures are to the glottis, the higher they are above 500 cps.

At this point, we should repeat that consonants are not perceived by steady-state frequencies of the formants, as are vowels, but by rapidly changing frequencies reflecting articulatory movements. It is for simplicity's sake that we mention only one frequency for the first formant of /Ṛ/ or for /ḉ/ -- their frequency at the instant of maximal stricture (550 cps for /Ṛ/, 1000 cps for /ḉ/). Acoustically, the first formants of such consonants are not well described by one frequency but by rapidly rising or rapidly falling formant transitions. (The frequency toward which formant transitions move for the perception of consonants is called their <u>locus</u>: 550 cps and 1000 cps are the first-formant <u>loci</u> of /Ṛ/ and /ḉ/.) For instance, knowing that the first formants of Arabic vowels vary from about 250 cps for /i/ and /u/ to 750 cps for /a/, we can say that, for the perception of /ḉ/, first-formant transitions rise (toward 1000 cps) after all vowels, but that, for the perception of /Ṛ/, first-formant transitions rise only after /i/ and /u/ (from 250 cps to 550 cps) and fall after /a/ (from 750 cps to 550 cps). All this is confirmed by the analysis of spectrograms and verified by synthesis.

Let us return to the x-ray configurations of Row 3. The articulatory production of /Ṛ/ should be described as a circular motion of the tongue root, the bulge moving first horizontally toward the pharyngeal wall, then vertically along that wall. A similar circular motion occurs in other languages, as will be shown in the following pages.

Row 4, left, shows that the voiceless constrictive /χ/ may well be regarded as the voiceless counterpart of the voiced /Ṛ/, as generally assumed. Its pharyngeal stricture is about at the same high level along the pharyngeal wall, and it is also produced in a circular movement of the tongue root, first horizontally toward the pharynx, then rising along the pharyngeal wall. The volume of the pharyngeal cavity and the first-formant frequency of /χ/ are also similar to those of /Ṛ/. The differences are simply those which one would expect to find in a voiceless constrictive: the stricture between the tongue bulge and the pharyngeal wall is narrower for /χ/ than for /Ṛ/, and the uvula does not curl into a trill position but lies flat over the tongue bulge to prolong the stricture and contribute to the production of a friction turbulence. Even in this flat position, however, the uvula may (but does not always) move intermittently toward and away from the tongue in order to produce slight variations in the air-stream output, variations which enhance the perceptual effect of the friction.

But it must be emphasized that the perception of /χ/, like that of /Ṛ/, does not depend so much upon the friction sound produced during maximal stricture as upon the formant-transitions[1] sound produced, during the tongue-root motions toward and away from the pharyngeal wall, by changes in the volume of the pharyngeal and mouth cavities. The all-important role played by formant transitions in the perception of consonants has been well demonstrated by synthesis.

Row 5, left, shows the mouth configuration of a fifth pharyngeal consonant of Arabic, the /q/. This consonant is not a constrictive, as the other four consonants in Fig. 1, but a voiceless plosive which contrasts, in Arabic, with the velar plosive /k/, of which it is considered the emphatic counterpart. Except for the complete interruption of the air-stream and the transition features which characterize all voiceless plosives, this arabic /q/ presents the same kind of pharyngeal acoustic features as the constrictive /χ/.

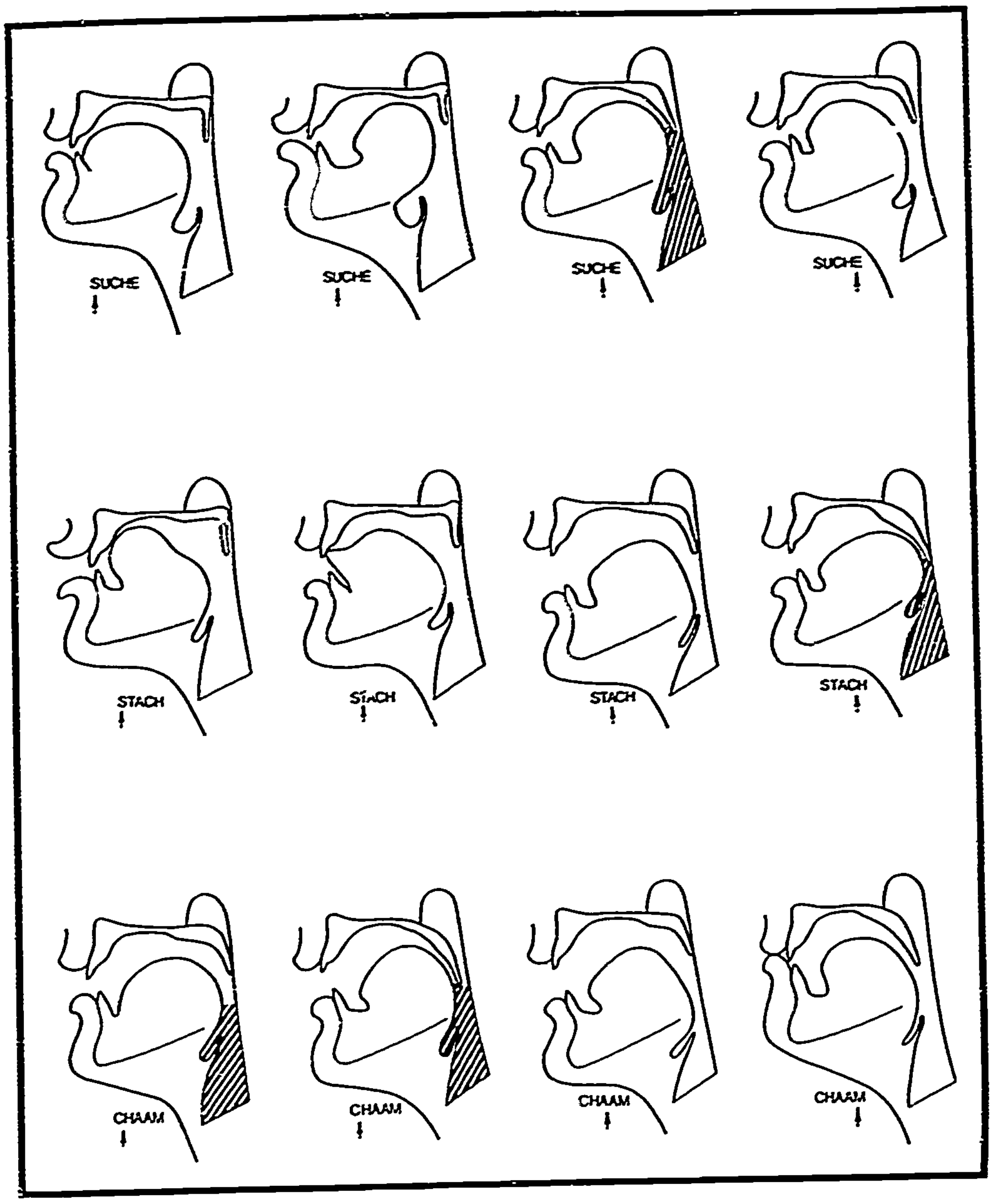
In the rest of this study, the four Arabic constrictives of Fig. 1, and the /a/ vowel will serve as references for the description of pharyngeal features in three German, one Spanish, one French, and one American sound.

GERMAN

Three figures are devoted to the pharyngeal features of German, one for the voiceless constrictive /x/, and two for the voiced constrictive /r/. A third figure is necessary because in postvocalic, word-final position the German /r/ is much more vocalic than in all the other positions. To distinguish the two allophones, we shall transcribe the final variety as /-r/.

Fig. 2 presents the German "Achlaut," /x/, in medial, final, and initial position, adjacent to back and central

GERMAN



INTERVOCALIC, FINAL, AND INITIAL ACHLAUT

Figure 2

vowels, in the words Suche /zuxə/, stach /ʃtax/, and Chaam /xɑm/. In each one of those three positions it is comparable to the Arabic /χ/, but with slightly reduced tongue backing and slightly wider strictures between the tongue and the pharyngeal wall in the initial stage (horizontal withdrawal of the tongue bulge toward the pharynx). These differences might justify the use of the phonetic symbol /x/ rather than /χ/.

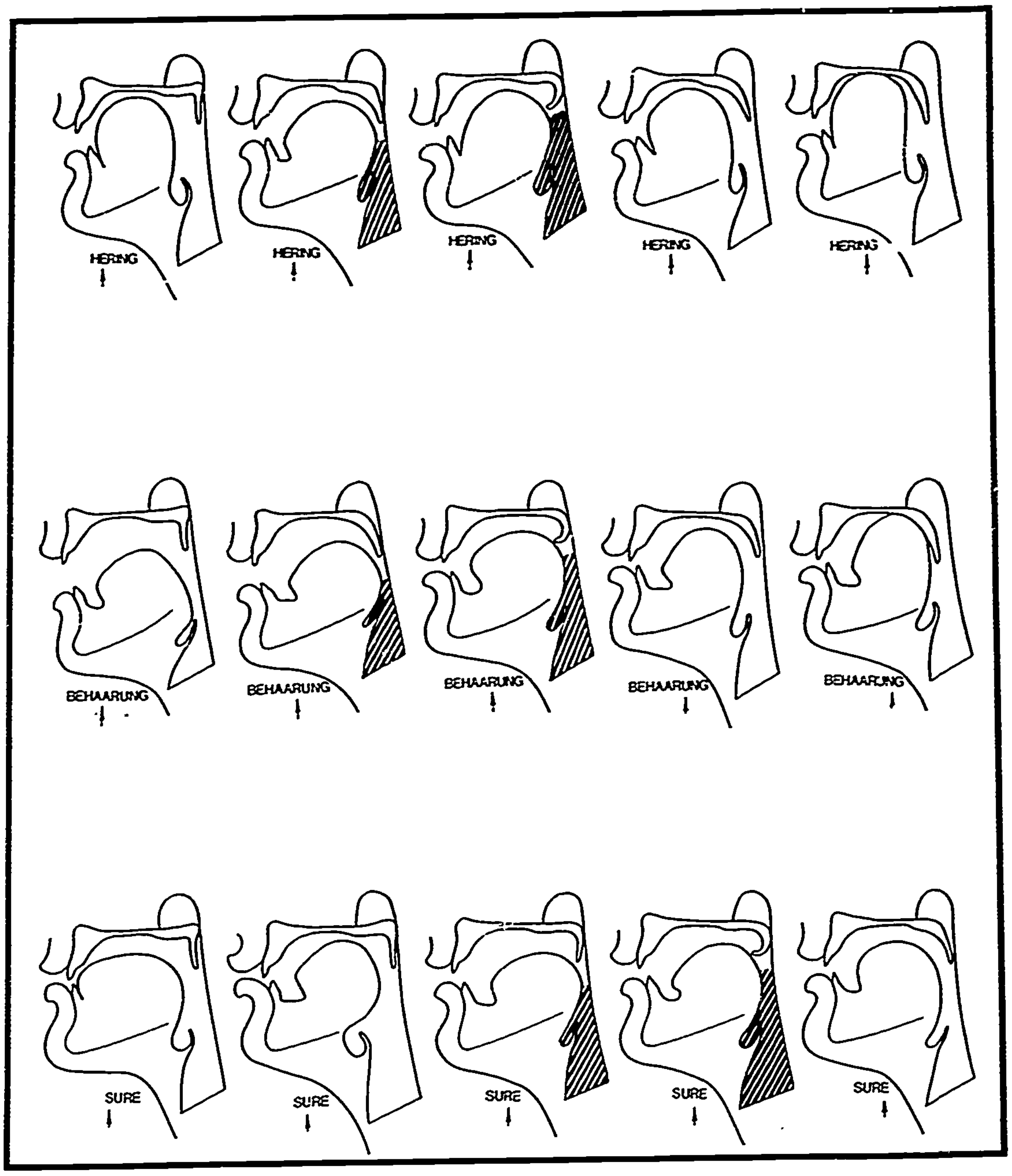
In Suche, pharyngealization is indicated by the fact that the tongue (which is very high for /u/) does more than just withdraw straight back, it also lowers considerably toward the pharyngeal wall to produce a stricture in the high pharynx area -- the very area that characterizes the Arabic /χ/. The main factor in the perception of this Achlaut is not the friction noise that occurs between the tongue and the extreme end of the velum but the rise in the first-formant transition which is produced by the backing of the tongue root to reduce sharply the volume of the pharyngeal cavity. In stach, on the contrary, the volume of the pharyngeal cavity increases from /a/ to /x/ while the tongue bulge rises along the pharyngeal wall. This is precisely the same kind of motion as in Arabic /χa/, but in reverse. And in Chaam /xɑm/, Fig. 2, Row 3, the analogy with Arabic /χa/ (Fig. 1, Row 4) is complete -- before rising to a high-pharynx area in the second frame the tongue draws back to form a mid-pharynx stricture which begins the circling motion of the Arabic /ʁ̞/ and /χ/. Only after the back-and-rise circling motion is completed does the tongue bulge start lowering toward the low-pharynx area of /a/.

A more complete sequence of x-ray frames for Chaam is shown in Fig. 11 (Row 3), a figure devoted to comparing the circling motion in three languages. There it appears more clearly that the circling motion is a back-and-rising one, for after the tongue bulge has drawn back to mid pharynx, it rises for two frames along the extremity of the velum before starting its descent toward the low pharynx for /a/.

Fig. 3 presents the standard /r/ of Northern German, that is, the variety of /r/ which does not involve the tip of the tongue but involves the back and root of it. This German /r/ is shown in intervocalic position after a front, a central, and a back vowel in the words Hering, Behaarung, and Sure. (This last word allows a minimal comparison with Suche in Fig. 2.) In each case two /r/-frames are selected in order to give an idea of the circling motion of the tongue, which is so striking when the film is shown at 24 frames per second.

In Hering we see the tongue draw back toward mid pharynx (Frame 2), sharply reducing the volume of the pharyngeal cavity and producing a rise in the first-formant frequency. Simultaneously, an enlargement of the mouth cavity produces a sharp fall in the second formant so that the two formants come close together. Then the tongue bulge rises along the pharyngeal wall (Frame 3), and the uvula joins the tongue just above the stricture to produce some loud trills which appear on the spectrograms as sharp periodic interruptions of the air-stream, recurring from two to six times. Finally the uninterrupted circling motion of the tongue places it out of reach of the uvula on its way to the next vowel. In

GERMAN



INTERVOCALIC r

Figure 3

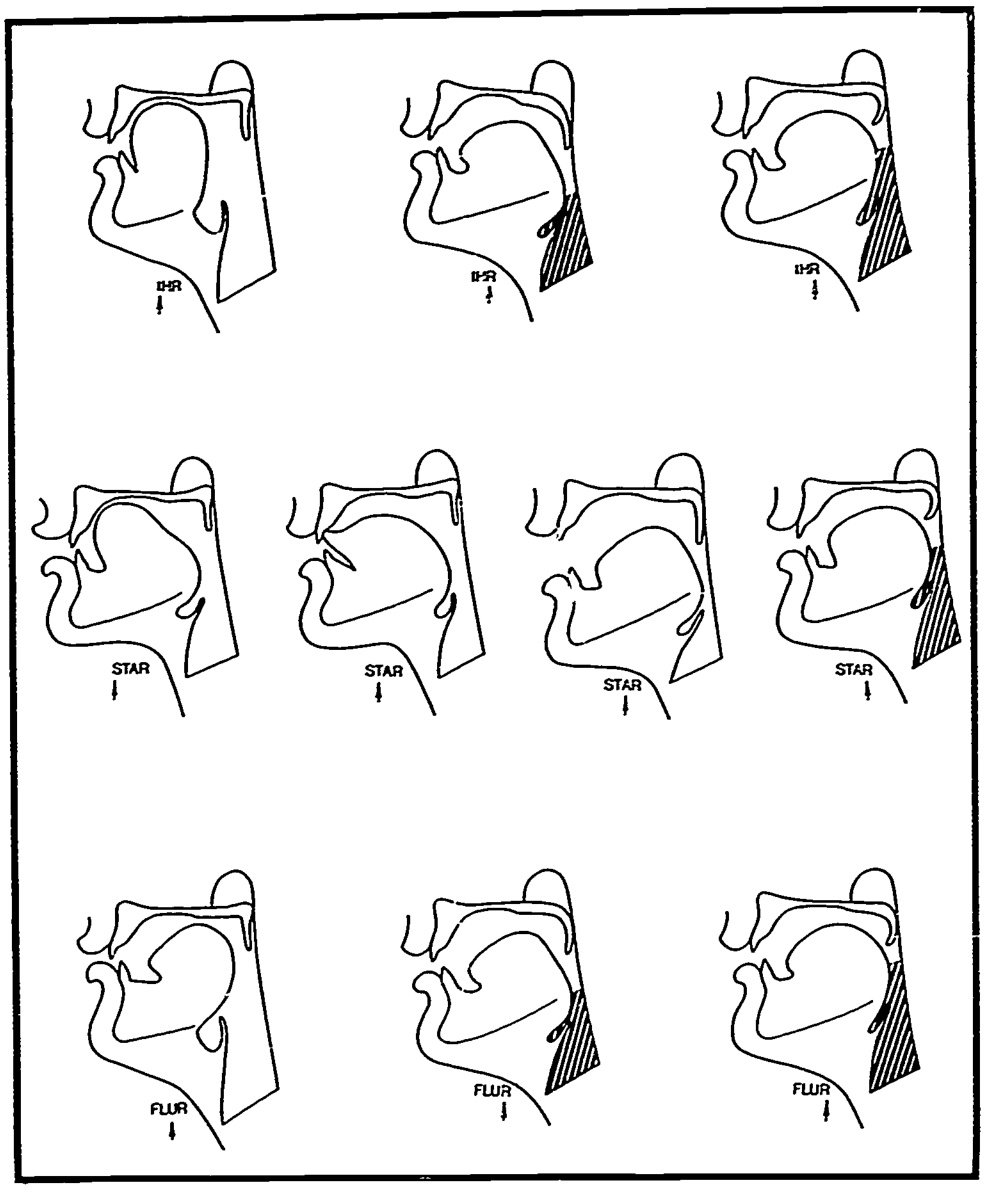
Behaarung, we see the tongue-root bulge, which is already formed for /a/ (Frame 1), rise along the pharyngeal wall to mid pharynx while narrowing the stricture (Frame 2), then rise further to the level of high pharynx (Frame 3) so as to permit the uvula to trill loudly. In Sure the sequence is about the same as in Hering, but the circling motion of the tongue is even clearer. (This motion is more pronounced than in Suche of Fig. 2.) Synthesis shows that, in all three words, the /r/ sound has already been perceived when the tongue bulge reaches mid pharynx, and that it is only enhanced by the uvular trills that follow. We can therefore assume that, in the perception of German /r/, the rise of the first-formant transition is the primary cue, and the uvular trills, loud as they are, must be considered secondary.

Fig. 4 presents the articulation of the German voiced constrictive /-r/ in final position, after a front, a central, and a back vowel in the words ihr, Star, and Flur. (What will be said applies only to the "weak" final /-r/, not to the "strong" final /-r/ of words like irr, Herr, starr which behaves like medial or initial /r/.)

In final position, the circling motion of /-r/ is much more extended than in other positions, and the uvular action much less. The circling motion is so extended that the tongue comes close to going through an /a/ position. But the sound of /a/ is obscured by the fact that the jaws do not separate as for a clear /a/ vowel.

In ihr (Row 1), the tongue root bulges back toward the lower pharynx to form a stricture no more than one or two

GERMAN



FINAL r

Figure 4

centimeters above the typical place for an /a/ stricture. Then the tongue bulge rises (but very slowly) along the pharyngeal wall and the sound of /-r/ dies out as the bulge lightly contacts the uvula. During all that time the jaws fail to open. In Flur (Row 3) the action of the tongue is quite similar and the jaws also fail to open. In Star (Row 2) the tongue simply rises slowly along the pharyngeal wall (Frame 4) and the jaws close noticeably. (After /a/ the sound of /-r/ is less audible than after /i/ or /u/.)

The frames of Figs. 2, 3, and 4 are selected from, and representative of, four films in which German natives recorded more than 50 words each. Except after /a/ and /ə/ the final /-r/ always glided through an obscure /a/ and ended with a very light friction sound.

To learn more about those /-r/'s we played the film in reverse for three American visitors one day, and in every case they heard an /a/ in the reversed glide; but this /a/ did not terminate the reversed glide -- it was followed by either an [x] or an [R] sound. For instance, for reversed Flur, they repeated either [Raul] or [xaul], for reversed wir, they repeated either [Raiv] or [xaiv], etc. The relation between final /-r/ and the vowel /a/ is therefore not an illusion, but this /a/ is generally obscure and is followed by a light pharyngeal constriction as the syllable ends.

SPANISH

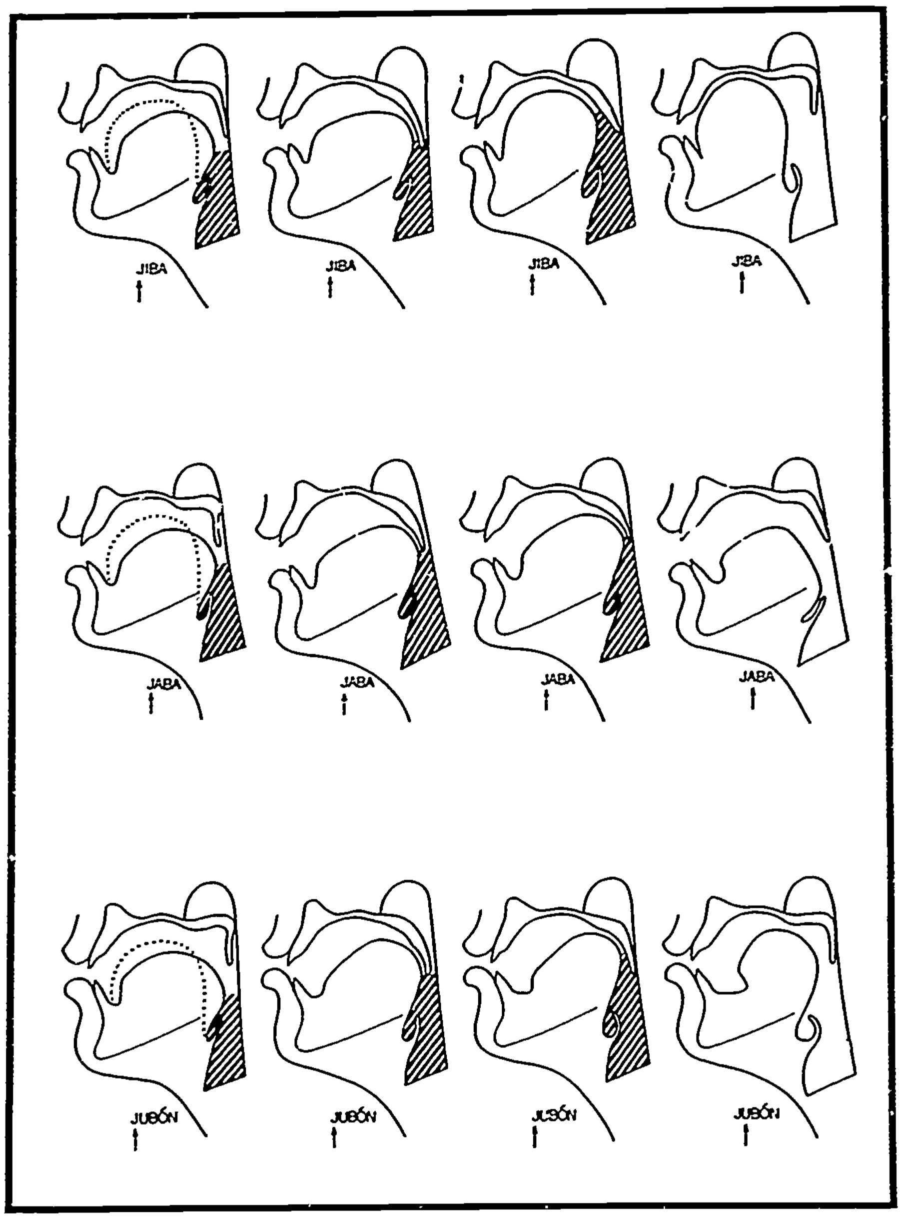
Figs. 5 and 6 present, in every row, two key configurations for the articulation of the Spanish "jota," /x/, a voiceless constrictive, as spoken by a cultivated native of Madrid at a rather fast syllabic rate.

In Fig. 5, the jota is in initial position, followed by a front, a central, and a back vowel in the words Jiba, Jaba, Jubón. In the first frames of each row, the tongue shape for the neutral vowel from which the tongue starts its /x/ motion is marked in dotted contour. It can be seen in the first row of Fig. 5 that, starting from this neutral position, the tongue first withdraws straight back toward the pharyngeal wall to form a mid-pharynx stricture, then the constriction rises to reach the extreme end of the velum and moves forward attracted by the high-fronted /i/ position. In the second row of Fig. 5, the tongue root starts again by moving straight back to the pharyngeal wall, then the stricture rises to the extreme end of the velum (Frames 2 and 3) and moves downward to the /a/ position in the lower pharynx. In the third row of Fig. 5, the tongue motion is comparable to that of the first row: backing, rising, and fronting to the high palato-velar position of /u/.

In all three rows, we can observe a back-and-rise circling motion of the tongue. With /i/ and /u/ this motion is continued by fronting, but with /a/ it is continued by falling. What is common to all three rows is the back-and-rise motion. After the rise which permits the friction noise between the tongue and the extreme end of the velum the tongue moves toward the next vowel position, in whichever direction that may be.

Fig. 6 presents the jota in intervocalic position after /i/, /a/, and /u/ in the words Lija, Raja, and Grújalo. In

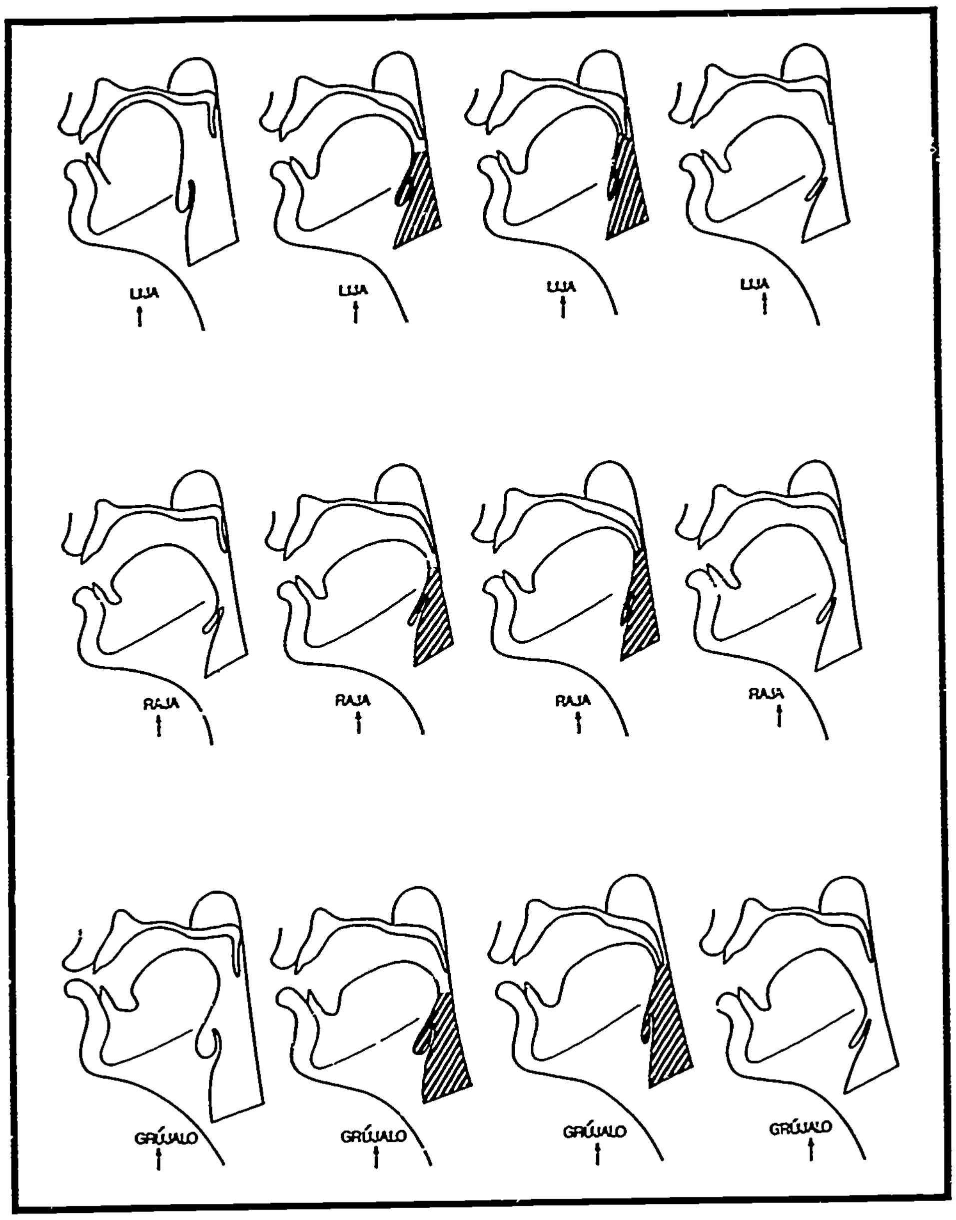
SPANISH



JOTA IN INITIAL POSITION

Figure 5

SPANISH



JOTA IN MEDIAL POSITION

Figure 6

each row, as in Fig. 5, the tongue tends first to form a stricture at mid pharynx, after which the constriction rises to the extreme end of the velum.

The circling motion of the tongue can be observed in more detail in Fig. 11, Row 2, where a complete sequence of frames is presented for the word Déjese. In Frames 3 and 4, the tongue bulge is drawn straight back toward mid pharynx, in Frames 5 and 6 it rises along the extremity of the velum, and at Frame 7 it has left the velum to move toward the /e/ position of Frame 8.

Acoustically, after /i/ and /u/ the withdrawal of the tongue toward the pharynx reduces the volume of the pharyngeal cavity and causes the first-formant transition to rise sharply. From /a/ to /x/, the volume of the pharyngeal cavity increases slightly, and the first-formant transition falls. After all vowels except /u/ and /o/ the volume of the mouth cavity increases and the second formant falls. After /u/ and /o/ the second formant is already too low to move lower.

Let us now compare Figs. 5 and 6 with the Arabic sounds of Fig. 1.

The Spanish jota is obviously comparable to the Arabic /χ/ of Row 4, in Fig. 1. Both are voiceless, both form their maximal stricture between the tongue root and the extremity of the velum, and both arrive at that constrictive position through a circling motion which first brings the tongue root to mid pharynx. The correlations between cavity modifications and formant transitions are also comparable.

Naturally, one would expect the jota to be similar to the Arabic /χ/ since it was introduced into the Spanish language by the Arab invaders who remained in Spain from the 7th to the 14th centuries -- a classic illustration of superstratum influence.

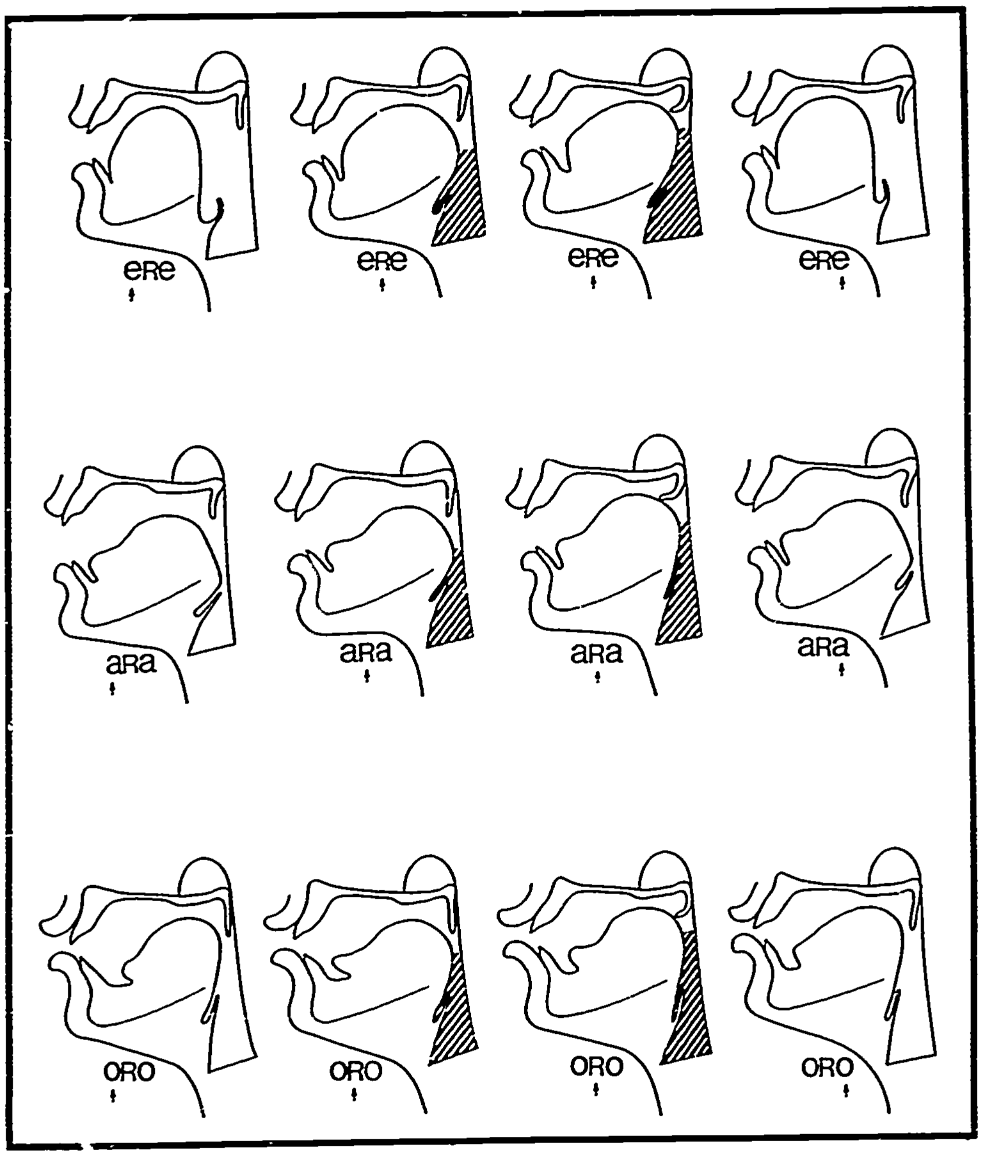
FRENCH

Fig. 7 presents the French /R/, a voiced constrictive, in medial position between front vowels in /eRe/, central vowels in /aRa/, and back vowels in /oRo/.

In /eRe/, the tongue root first draws back toward mid pharynx, a little higher than the level of an /a/ bulge (Frame 2), then the tongue bulge rises along the pharyngeal wall to permit the uvula to come lightly into contact with the tongue and to be raised intermittently by the air-stream pressure just above the stricture, and it continues forward toward the high-front position of the following /e/. The total motion of the tongue is a circling one: lowering, backing, rising, and fronting.

To observe this circling motion of the tongue in detail, we have traced in Fig. 11, Row 1, a more complete /yRy/ sequence, pronounced by another French speaker. There, the characteristic backing and rising motions of the tongue appear clearly. Starting from a high-fronted /y/ position (Frame 1) the tongue lowers and draws back toward the pharyngeal wall (Frames 2, 3, and 4) and rises along this wall (Frames 5, 6, and 7) before moving toward the next vowel (Frames 8, 9, and 10) -- in this case it means moving forward, but in other cases it could mean moving in some

FRENCH



INTERVOCALIC R

Figure 7

other direction (downward, for instance, if the next vowel were an /a/).

To understand the acoustic factors (the rapid musical changes) which explain the perception of French R, we must look at the articulatory configurations of /yRy/ in terms of cavity volumes and of the resulting notes of resonance.

When the tongue root withdraws toward the pharynx, the volume of the pharyngeal cavity (below the pharyngeal stricture) reduces sharply and the volume of the mouth cavity (behind the lips) increases sharply. The decrease of the pharyngeal cavity causes the first-formant frequency to rise, and the increase of the mouth cavity causes the second-formant frequency to fall, so that the two formants come close together. It is these simultaneous changes -- the rise of the first formant and the fall of the second formant to certain frequencies (their loci) -- which account for the perception of /R/ rather than the intermittent interruptions of the sound caused by the uvular trills.

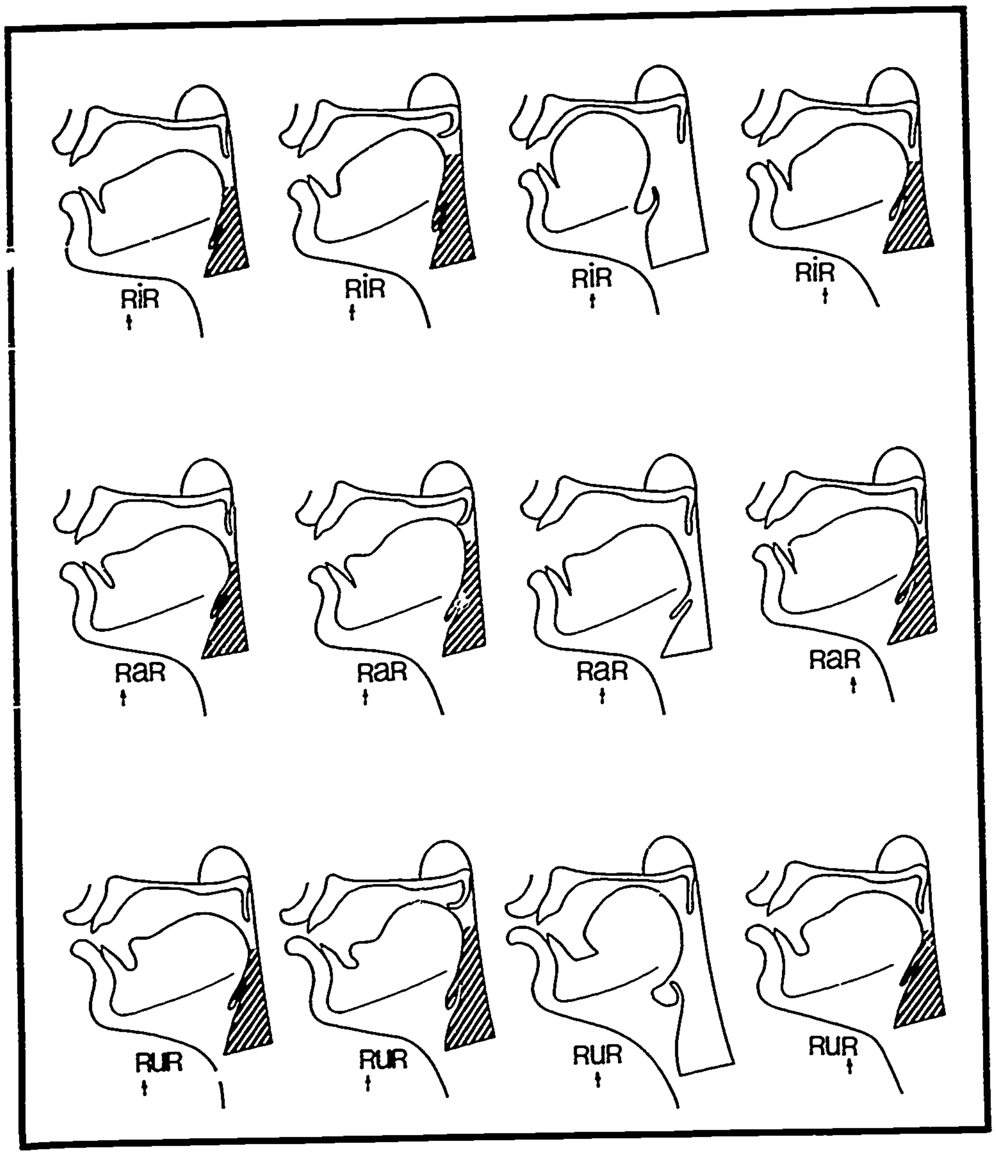
Let us return to Fig. 7. Since Row 2 begins with an /a/, the tongue root is already considerably withdrawn toward the pharynx. We see the tongue bulge draw back a little more (a consonantal stricture is narrower than a vocalic one) and rise along the pharyngeal wall until it is high enough to permit uvular trills to be produced. Then the tongue bulge falls rapidly toward the low pharynx level of the /a/ bulge. In terms of cavities, as the tongue bulge rises, the volume of the pharyngeal cavity increases to that of an /R/, and the first-formant frequency lowers from the /a/ level of about 750 cps to the /R/ level of about 550 cps.

We see that, whereas in the cases of /yR/ and /eR/ the first formant <u>rises</u> to the /R/ level, in the case of /aR/ the first formant <u>falls</u> to the /R/ level. /R/ first-formants, therefore, have their <u>locus</u> -- a frequency toward which formants move -- situated below the level of the /a/ first-formant, but above the level of the first formant for /y/, /e/, and all other vowels but /a/.

In /oRo/, Row 3 of Fig. 7, the circular movement of the tongue is evident, although less marked than in /iRi/ or /yRy/. The tongue bulge lowers and draws back toward the pharynx (Frame 2), then rises along the pharyngeal wall and moves forward to the /o/ stricture. From /o/ to /R/, the pharyngeal-cavity volume decreases and first-formant frequency rises.

Fig. 8 presents the same French /R/ as in Fig. 7, but in different syllabic positions -- in initial and final positions -- to see whether the /R/ characteristics observed for medial /R/ in Fig. 7 are preserved in the other positions. Fig. 8 shows that, in initial position, the circular (back and up) motion of the tongue is perhaps more marked, and the uvular trill contribution slightly stronger, than in medial position. In final position, on the contrary, the circular motion is less complete, and the uvular contribution generally by-passed. Nevertheless, the tongue withdrawal toward the pharynx is very sharp (notice the extended tongue motions which take place between Frames 3 and 4), but after the tongue backing, the bulge does not rise high enough to allow

FRENCH



INITIAL AND FINAL R

Figure 8

(or incite) any action of the uvula. It is, therefore, possible for French ears to perceive an /R/ very well without the contribution of uvular trills.

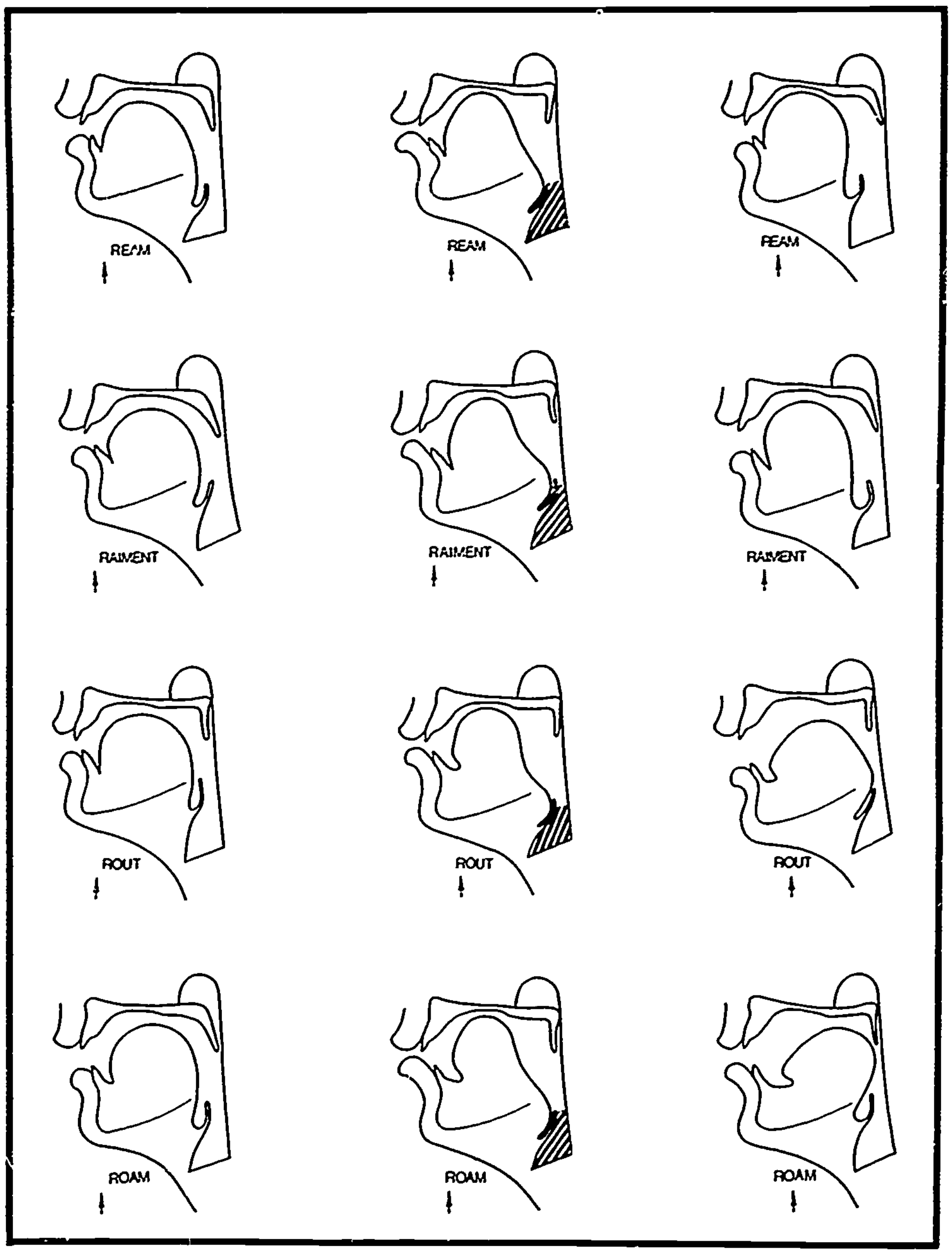
This last remark led us to experiment with the French /R/ by synthetic manipulation. It was found that the primary factor for the perception of /R/ is the high locus of the first-formant transition (which rises after all vowels but /a/) and the low locus of the second-formant transition (which falls after most vowels). The intermittent intensity variations of the air-stream (not really "interruptions" in Northern French) produced by uvular trills are secondary or superfluous, but the more they are used the greater is the perceptibility of the /R/; they give emphasis to a syllable.

A comparison of Figs. 7 and 8 with Fig. 1, Row 3, shows considerable similarity between the French /R/ and the Arabic /Ṛ/. Both are voiced, both reach their maximal degree of constriction in the high pharynx, both use light, intermittent uvular trills with moderation, and both are perceived primarily by a rise in the first-formant transition, (except after /a/) and a fall in the second-formant transition, which correlate with a volume decrease in the pharyngeal cavity (except after /a/) and a volume increase in the mouth cavity.

AMERICAN ENGLISH

Figs. 9 and 10 present the characteristic articulatory configuration of American /r/ at its moment of maximal constriction, in initial, medial, and final positions, in the words <u>Ream</u>, <u>Raiment</u>, <u>Rout</u>, <u>Roam</u>, <u>Mirror</u>, <u>Borrow</u>, <u>Coral</u>,

AMERICAN ENGLISH



INITIAL r

Figure 9

Later. A variety of tongue shapes can produce the American /r/; the one presented here is by far the most typical, according to an extensive study we have completed ("A dialect study of American r's by x-ray motion picture," to appear in Linguistics). But whatever the tongue shape -- apical, retroflexed, laminal, dorsal, or bunched -- if the American /r/ has that particular barking sound which correlates with a sharp lowering of the third formant, it shows two tongue strictures, one at the palate and one at the pharynx.

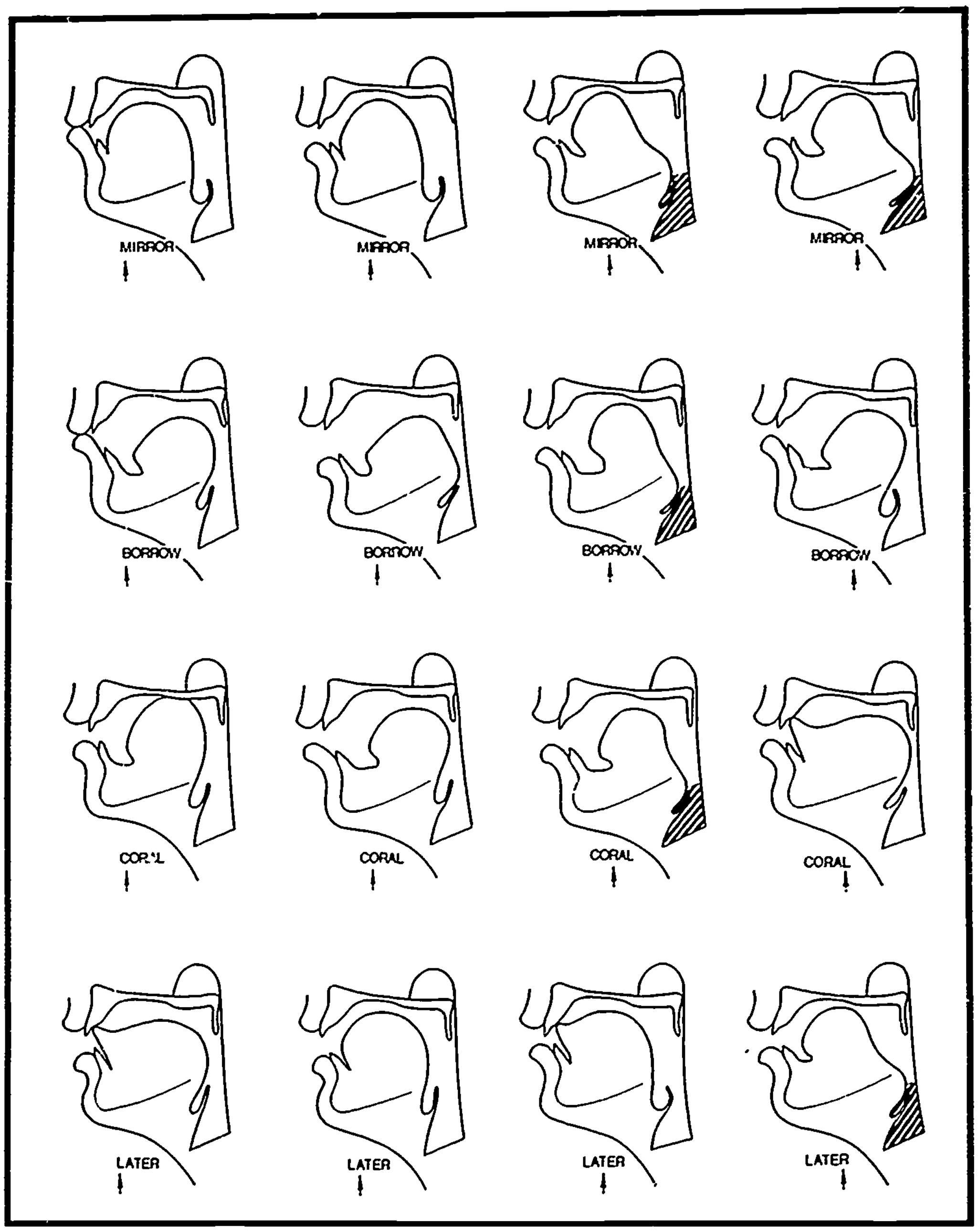
It is the pharyngeal stricture of American /r/ which prompts us to include it in this general study of pharyngealization.

In Fig. 9, the neutral position taken by the tongue at rest, just before the /r/ motion, is shown (first frame in each row) to help the reader realize the double direction of that motion. The tongue moves simultaneously toward the palate and the pharynx. The dorsum rises toward mid palate, and the root draws back toward the low pharyngeal wall, while the tongue back assumes a cambered shape remindful of a camel's back.

Figure 10 shows the same double bulge for /r/ in medial position (Mirror, Borrow, Coral), but slightly less marked than in initial position. It also suggests that the sharpest bulging toward the pharyngeal wall occurs in final position (Mirror, Later) -- the final /r/ of Mirror is considerably more backed than the medial one, and it is similar to the final /r/ of Later.

The double-bulge shape of American /r/ is strikingly

AMERICAN ENGLISH



INTERVOCALIC AND FINAL r

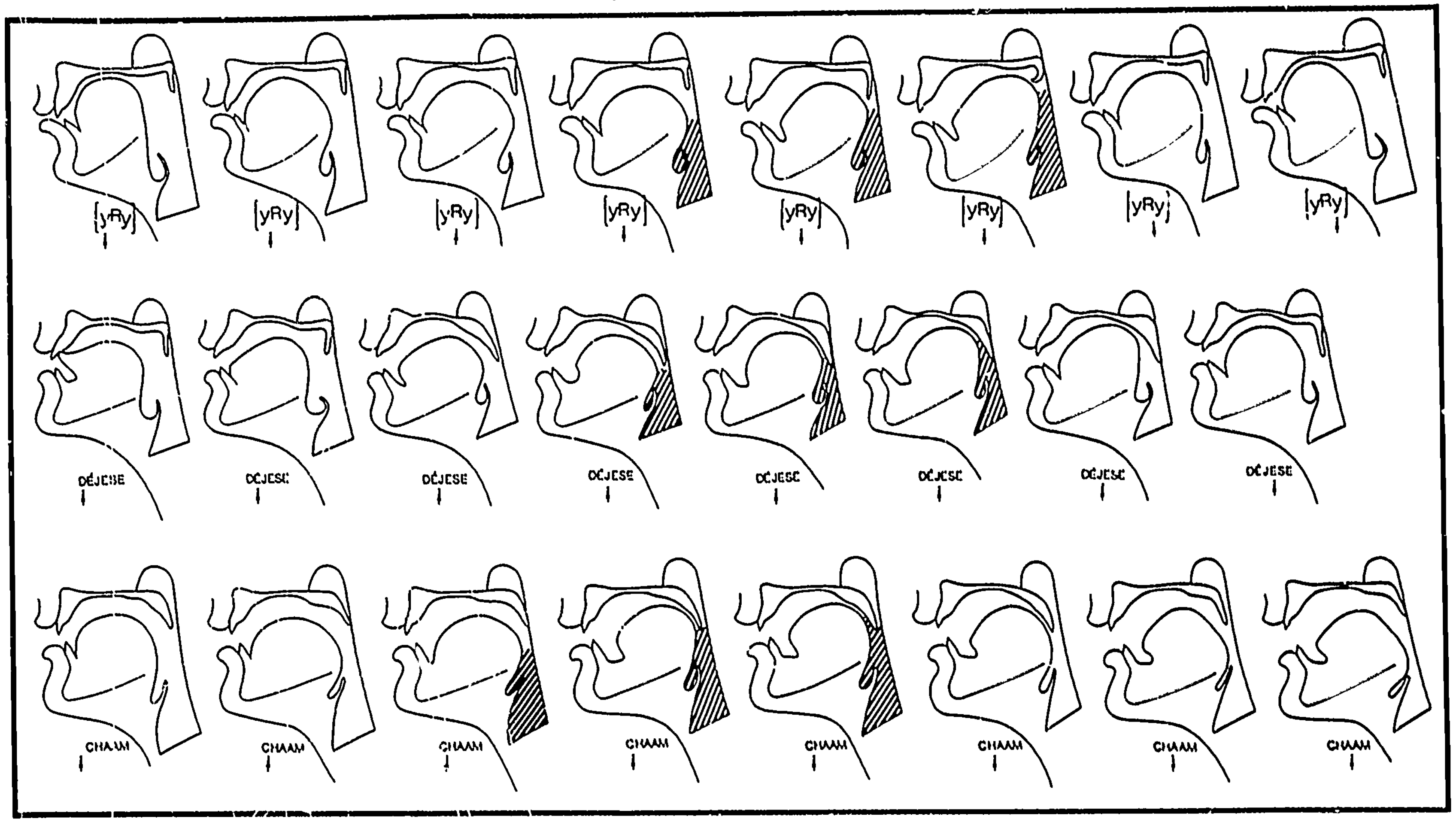
Figure 10

remindful of the tongue shape of Arabic /ḥ/ (Fig. 1, Row 2). It is true that important differences appear in the narrowness of the strictures -- for the Arabic /ḥ/ the radical stricture is narrower and the dorsal stricture is much wider -- but the general shapes are similar. The main analogy is in the radical bulge. The American-/r/ bulge is perhaps not quite so low as the Arabic-/ḥ/ bulge, but it is just about as low as the Arabic-/ç/ bulge, which compares better since it belongs to a voiced sound.

This comparison between an American and an Arabic sound raises a question of acoustic theory. Since the low level of the radico-pharyngeal bulge and the smallness of the pharyngeal cavity it creates are responsible for the high first-formant locus of /ç/ and /ḥ/, why is the first-formant locus of American /r/ much lower than that of /ç/? The answer must be in the narrowness of the <u>palatal</u> stricture of /r/. Just as labial rounding, acting as a second stricture of the whole tube, lowers the first-formant frequency of the vowels that have a back stricture, like /o/ or /u/, so does the dorso-palatal stricture of /r/ acting as a second stricture of the whole tube bring the first-formant frequency of the vocal tract lower than for a single stricture in the pharynx.

Fig. 11 presents the circling motion of the tongue for high pharyngeal constrictives in more complete sequences than was permitted by space in the French, Spanish, and German figures. These circling sequences have been mentioned at an appropriate place for each language and need not be repeated here.

FRENCH, SPANISH, AND GERMAN



THE CIRCLING MOTION OF PHARYNGEAL CONSTRICTIONS
Figure 11

SUMMARY

The development of the 9-inch image intensifier, which now makes it possible to observe the posterior regions of the vocal tract at the same time as the anterior ones, in cineradiography, has revealed that many speech sounds heretofore classed as velars are perhaps more pharyngeal than velar. In order to evaluate the pharyngeal quality of such "velar" consonants of German, Spanish, French, and English in articulatory, acoustic, and perceptual terms, the speech sounds that are most clearly pharyngeal are used as a reference.

The first part of this study is a description of the pharyngeal references: the /a/ vowel, with its stricture in the lower pharynx, and the five pharyngeal consonants of Arabic, two of them, /ç/ and /ḥ/, with a stricture in the lower pharynx, below the /a/ stricture, and the three others, /Ṛ/, /χ/, and /q/, with a stricture in the upper pharynx, above the /a/ stricture. The second part is a comparison of certain German, Spanish, French, and American English articulations with the vowel /a/ and the Arabic pharyngeals.

It is found that the German "Achlaut" /x/ and the Spanish "jota" /x/ are comparable to the Arabic /χ/, as voiceless constrictives of the high pharynx. The German /r/ and the French /R/ are comparable to the Arabic /Ṛ/, as voiced constrictives of the high pharynx, except that the uvular trills are much stronger, more regular, and more periodic in the German consonant than in the French and Arabic ones. The final /-r/ of German is not included in this statement, however -- it is equally comparable to

Arabic /Ṛ/ and Arabic /ç/, its stricture being at mid pharynx between the high-pharynx stricture of /Ṛ/ and the low-pharynx stricture of /ç/. Finally, the low pharyngeal stricture of the American /r/ is comparable to that of Arabic /ḥ/ or /ç/. The analogy between these Arabic and American sounds is also visible in the double-bulge shape of the whole tongue which is found in both the American /r/ and the Arabic /ḥ/.

The backing-and-rising "cirlcing" motion of the Arabic high pharyngeals is clearly present in the German, Spanish, and French high pharyngeals. Acoustically, this last analogy is reflected by a similar rise of the first formant and fall of the second formant after vowels other than /a/, rise and fall which form the basis of their auditory perception.

This study is based on x-ray motion pictures, and on spectrographic analysis verified by artificial-speech synthesis. It includes ten figures which show 34 articulatory sequences of x-ray frames selected from motion pictures taken at 24 frames per second.

CONSONANT GEMINATION IN FOUR LANGUAGES:

AN ACOUSTIC, PERCEPTUAL, AND RADIOGRAPHIC STUDY

'Gemination' applies here to the meaningful perceptual doubling of a consonant phoneme. It occurs frequently across word boundary, as in English Will lend vs. Will end, in German Stiehl Loden vs. Stiehl Oden, in Spanish El lecho vs. El hecho, and in French Il l'aime vs. Il aime. It also occurs, but less generally, within word boundary, as in Spanish Perro vs. Pero, in French Il acquerrait vs. Il acquérait, or Il serrerait vs. Il serrait, and in German starr vs. Star, Beharrung vs. Behaarung. (The term 'double' -- not to be confused with geminate -- is usually reserved for graphic symbolization of two consonants.) In the preceding examples a difference of vowel color, or of vowel length may occur concomitantly -- the /e/'s and /a/'s are shorter and less fronted before geminate consonants than before single ones -- but the gemination always seems to make a major contribution to the distinction of meaning. That contribution could be more significant than above, as in Spanish Carro vs. Caro, in French Mourrait vs. Mourait; or it could be less significant, as in German wirr vs. wir, where the difference of vowel color and vowel length may be considered as playing the major role; but the linguistic function of gemination can never be denied. (Whether the vowel of wirr conditions the gemination of the final consonant or whether, on the contrary, the consonant gemination conditions the length and color of the preceding vowel, is a matter of conjecture.)

The object of this study is to examine the acoustic and the articulatory correlates of consonant gemination, both across and within word boundary, and to compare their behavior among four languages -- English, German, Spanish, and French. The technical procedures will be somewhat similar in all cases and for all languages, and can be described once and for all.

A. Appropriate utterances are composed in each language, making maximal use of minimal pairs (German Nenn Omen vs. Nenn Nomen) or near minimal pairs (German schnorren vs. schmoren, Spanish Un nicle vs. Unible) in order that the variables be in similar phonetic environment.

B. The utterances are recorded by a few native speakers of each language, and spectrograms are made on which it is possible to measure consonant duration in centiseconds (cs) and observe the tempo of formant transitions and the variations of overall amplitude. The duration of adjacent vowels is also of interest in certain cases. Consonant durations are averaged for geminate and single consonants in each position.

C. Motion picture x-rays of selected utterances are made, in our cineradiographic studio, with correlated sound, again using native speakers of each language. These films offer a profile view of the motions of the tongue, the jaws, the velum, and the lips, and show the places of constriction inside the mouth and the continuous modifications of the mouth cavities, in shape and volume, during the speech process.

D. The articulatory movements of geminate consonants and single consonants are studied concurrently on x-ray motion pictures (at normal speed and in slow motion) and on spectrograms. Significant sequences of tongue movements are traced from x-ray frames by means of appropriate enlargers in order to make the comparison between geminate and single consonants easily accessible to the eye.

E. Finally, artificial-speech synthesizers are used to test by ear what consonant durations, or ratios of consonant durations are appropriate in each language in distinguishing geminate consonants from single ones. For example, the sentence I've seen Elly is synthesized ten times with ten different durations of the /n/ hold varying from 2 to 20 centiseconds in steps of 2 cs. Each one of the ten samples is recorded 5 times and the 50 recorded items are mixed in random order by tape splicing. Then several American listeners are asked to identify by ear each item as either I've seen Elly or I've seen Nelly. Analysis of the response data for each language indicates what durations are heard as single consonants and what as geminate consonants, and what range of durations is ambiguous. This sort of experiment may then be repeated for different classes of consonants to see whether ratios remain constant.

A. GEMINATION ACROSS WORD BOUNDARY

Only three consonants, /n/, /l/, and /s/, regularly occur in all four languages, both in final and initial position, in such a way as to permit the contrasting of geminate with single consonants across word boundary. And even then, we must overlook the fact that German /s/ voices to [z] in initial position. The consonant /r/ also occurs in the four languages in the appropriate positions and would produce distinctions in English (Her ace vs. Her race) and in French (Leur âge vs. Leur rage). But in German, /r/ lengthening in final position after short vowels would interfere with gemination: schnorr Eis might not be distinguishable from schnorr Reis unless a glottal stop occurred; and in Spanish, /r/ lengthening in initial position would practically neutralize such oppositions as Color rojo vs. Coló rojo (cf. Stockwell and Bowen, The Sounds of English and Spanish, p. 82). To facilitate comparison among the four languages, therefore, all our contrastive examples were chosen exclusively from the three consonants: /n/, /l/, and /s/. Fortunately, those three consonants do not represent a single category, but three very different ones: the nasals, the glides, and the constrictives.

For each one of these three phonemes /n/, /l/, and /s/, the geminate consonants will be compared to single consonants in two different syllabic positions: word final before a vowel and word initial after a vowel, and to the average of these two positions. In English, for instance, The race sends will be compared with The race ends and with The ray sends; in German, the Nenn Nomen will be compared with Nenn Omen, and Sie sahn Nomen with Er sah Nomen; in Spanish El lápiz will be compared with El Apis, and Es el lápiz with Ese lápiz; in French La masse sacrée will be compared with La masse agrée and with Le mât sacré.

1. THE DURATION FACTOR

Preliminary examinations of spectrograms have indicated that consonant duration is an important factor -- perhaps the

Table 1. ENGLISH

Geminate Consonant	Single Consonant		Average; Ratio of Geminate to Single Conson.
/n/			
I've seen Nelly 12.2 (7 ... 21)	I've seen Elly 7.1 (5 ... 10)	Final	7.9; 1.5 to 1
	We see Nelly 8.8 (7 ... 11)	Initial	
/l/			
It will lend 11.1 (7 ... 18)	It will end 7.9 (4 ... 11)	Final	8.5; 1.3 to 1
	And we lend 9.3 (8 ... 12)	Initial	
/s/			
The race sends 22.1 (20 ... 24)	The race ends 14.4 (11 ... 17)	Final	16.3; 1.3 to 1
	The ray sends 18.2 (16 ... 21)	Initial	

15.1 ← Total Average → 10.9

Total Ratio → 1.4 to 1

Table 2. GERMAN

Geminate Consonant	Single Consonant		Average; Ratio of Geminate to Single Conson.
/n/			
Beginn nimmer Nenn Nomen Ist das ein Nerz	Beginn immer Nenn Omen Ist das ein Erz 11.2 (8 ... 21)	Final	9.6; 1.5 to 1
Dann nahm er das Buch Sie stehn nah bei Sie sahn Nomen 14.5 (10 ... 22)	Da nahm er das Buch Sie steh nah bei Er sah Nomen 8.0 (6 ... 12)	Initial	
/l/			
Stiehl Loden Verzoll Leber So ist das Stilleben	Stiehl Oden Verzoll Eber Es ist still eben 11.5 (7 ... 19)	Final	10.1; 1.5 to 1
Schillenen Die Saal Lampen So ist das Schill Leben 14.8 (11 ... 20)	Schi Lehnen Sie sah Lampen So ist das Schi Leben 8.7 (6 ... 11)	Initial	
/s/			
Miss Seen	Missehen 14.5 (11 ... 22)	Final	13.5; 1.4 to 1
Das sah er 18.4 (16 ... 30)	Da sah er 7.2 (5 ... 9) [12.6]	Initial	
15.9 ←	Total Average		→ 10.7
	Total Ratio		→ 1.5 to 1

Table 3. SPANISH

Geminate Consonant	Single Consonant		Average; Ratio of Geminate to Single Conson.
/n/			
Un naire, Ven naves Es un nombre	Un aire, Ven aves Es un hombre 8.0 (6 ... 10)	Final	8.3; 2.1 to 1
Un nido, Un nicle 17.6 (14 ... 27)	Unido, Unible Unanime 8.6 (7 ... 10)	Initial	
/l/			
El lápiz, El lecho El limita, El loro	El Apis, El hecho El imita, El oro 9.4 (7 ... 14)	Final	9.4; 1.7 to 1
Es el lápiz 16.3 (14 ... 19)	Ese lápiz 9.5 (8 ... 10)	Initial	
/s/			
Las solas, Las salas Las sobras	Las olas, Las alas Las obras 8.6 (6 ... 13) [12.3]	Final	13.0; 1.5 to 1
Unos cincuenta Todos somos 19.7 (14 ... 22)	Uno cincuenta Todo somos 13.6 (10 ... 20)	Initial	

17.9 ← Total Average → 10.2

Total Ratio → 1.8 to 1

Table 4. FRENCH

Geminate Consonant	Single Consonant		Average; Ratio of Geminate to Single Conson.
/n/			
	L'une avale 8.2 (5 ... 10)	Final	
Une nasale 15.5 (13 ... 19)	L'u nasale 8.2 (6 ... 11)	Initial	8.2; 1.8 to 1
	Lunatique 6.9 (5 ... 8)		
/l/			
	La ville imite 7.6 (6 ... 9)	Final	
La ville limite 16.8 (12 ... 20)	La vie limite 7.4 (6 ... 8)	Initial	7.5; 2.2 to 1
	Les militants 6.0 (5 ... 7)		
/s/			
	La masse agrée 13.5 (11 ... 15)	Final	
La masse sacré 21 (17 ... 26)	Le mât sacré 13.4 (12 ... 15)	Initial	13.4; 1.6 to 1
	C'est massacré 12.1 (11 ... 13)		
17.8 ←	Total Average	→	9.7
	Total Ratio	→	1.9 to 1

most important, but certainly not the only one -- in the linguistic functioning of gemination. It is also the easiest factor to analyze, objectively. Therefore, we shall first examine the contribution of duration. Tables 1, 2, 3, and 4 have been drawn for this purpose. They present the complete list of words or utterances used in this analysis, with average durations in centiseconds for each of the three categories of consonants that are used. Let us now look at the statistical results of Tables 1 through 4, each table furnishing the duration data for a different language. The averages in columns 1 and 2 are based on two recordings by three, four, or five native speakers of each language. They represent, therefore, from 6 to 10 measurements per word or utterance. Following every figure representing an average, a parenthesis gives the shortest (left) and the longest (right) durations that were recorded for that category. In column 3 are shown the total averages for single consonants and the ratios of geminate consonants to single ones. At the bottom of each table are the total averages, as well as the duration ratio, of geminate to single consonants, based on the total averages.

In order to make the durations as comparable as possible among languages, some corrections were made in the German and the Spanish data. These corrections are indicated in brackets under the actual figures. They are necessitated by the /s/-voicing that occurs in German in initial position and in Spanish in final position when the next word begins with a vowel. For German, since all initial /s/'s following a word ending with a vowel are voiced (and are much shorter than an /s/ would normally be), we multiplied their real duration by a factor of 1.75 which corresponds to the minimal difference of duration between /s/ and /z/. Thus, the durations of initial /s/ in Da sah er is given in brackets as 12.6 cs rather than 7.2 cs. For Spanish, the final /s/ of Las olas, Las alas, and Las obras was voiced (and shorter than an /s/ would normally be) in nearly two-thirds of our recordings. To compensate for this voicing, the real duration was multiplied by a factor of 1.5; thus, the duration of Spanish final /s/ across word boundary is given in brackets as 12.3 cs rather than 8.6 cs.

The most general and obvious results are those of total averages and total ratios which appear at the bottom of each table. They offer contrasting figures which separate the two Latin languages from the two Germanic ones with respect to the role of duration in the meaningful opposition of geminate consonants to single ones. The duration ratio of geminate to single is clearly higher in the Latin languages (1.9 to 1 and 1.8 to 1) than in the Germanic languages (1.4 to 1 and 1.5 to 1). Furthermore, a comparison between the two Latin languages shows the ratio to be slightly higher in French (1.9 to 1) than in Spanish (1.8 to 1); and a comparison between the two Germanic languages shows the ratio to be higher in German (1.5 to 1) than in English (1.4 to 1). This can perhaps be interpreted as indicating that consonant gemination across word boundary is less distinct, less stressed, more slurred in English or German (especially in English) than in French or Spanish.

The duration differences between those two pairs of languages is visible at the level of single as well as of geminate consonants, although more at the latter level -- not only are the geminate shorter in English and German (15.1 and 15.9) than in French and Spanish (17.8 and 17.9), but the single consonants are also longer (10.9 and 10.7 in English and German vs. 9.7 and 10.2 in French and Spanish, respectively).

The manner in which the contrast is made at word boundary between geminate and single consonants may depend upon whether the single consonant is word-final (before a vowel) or word-initial (after a vowel). Let us, therefore, start by comparing the single consonants among themselves in each language.

With respect to duration differences between final consonants and initial consonants, French stands apart from the three other languages, although followed very closely by Spanish. In French (Table 4), final and initial consonants are practically of equal length. This is true of all three categories of consonants: final /n/, 8.2, initial /n/, 8.2; final /l/, 7.6, initial /l/, 7.4; final /s/, 13.5, initial /s/, 13.4. This equality partly confirms the belief of the first phoneticians that there is no auditory difference whatsoever between pairs like Un invalide vs. Un nain valide, Celui qu'il y voit vs. Celui qui lit voit, Les Russes ont fini vs. Les rues sont finies. Paul Passy, the originator of the phonetic alphabet (International Phonetic Association) had already said about French: "Word division has no effect upon syllable division. There is no difference between Les aunes and Les zones..." (Les Sons du Français, Paris, Didier, 1927, p. 61).

In Spanish (Table 3), the final consonants are shorter than the initial ones, but the differences appear nearly negligible, except in the case of /s/ (and even then it is not great if we use the corrected figure in the brackets): final /s/, 12.3, initial /s/, 13.6; final /l/, 9.4, initial /l/, 9.5; final /n/, 8.0, initial /n/, 8.6. Small as these differences are, they may reflect a tendency to slurr the final consonants more than the initial ones since, as we have mentioned, the first /s/ of Las olas, etc., tends to voice to [z].

In English, final consonants are clearly shorter than initial ones. This condition exists in all three categories: final /n/, 7.1, initial /n/, 8.8; final /l/, 7.9, initial /l/, 9.3; final /s/, 14.4, initial /s/, 18.2. This difference of duration is perhaps indicative of a tendency to slurr the final consonants more than the initial ones. This tendency is known to reach a high point when the final /t/ of utterances like Cut in, Let (h)er, At ease, and Beat it is nearly turned into a flap like the single apical tap of Spanish Caro.

This time-reduction of the final consonant before a vowel is perhaps also related to consonant anticipation, a marked characteristic of English which makes possible distinctions like Plain ice vs. Play nice, Sole aim vs. So lame, House ad vs. How sad. It is practically impossible for a Frenchman either to make or to understand such oppositions. There is little doubt that duration plays a part here, but on spectrograms it is mostly a matter of the distribution of intensity in the formant transitions which lead to and from the boundary consonant. In Sole aim, for instance, the arresting transitions of the /l/ would be strong and the minimum of intensity would follow them; whereas in So lame the arresting transitions of the /l/ would be weak, its releasing transitions would be strong, and the minimum of intensity would precede them.

In German, the duration conditions are the reverse of English. It is the final consonants that are the longer: final /n/, 11.2, initial /n/, 8.0; final /l/, 11.5, initial /l/, 8.7; final /s/, 14.5, initial /s/ (with correction), 12.6. This seems to be due to the glottal stop which occurs regularly when a final consonant is followed by an initial vowel. Instead of being slurred, as in English, the final consonant of German is slightly reinforced by the anticipation of the glottal closure, and a lengthening results. We asked our subjects to read the utterances like Beginn immer with a legato between /n/ and /ɪ/, but they could not do it. When asked to repeat in order to produce a legato some of them were able to do it, but those recordings sounded so unnatural that they had to be discarded. The spectrograms that were made of them showed that the forced-legato consonants were unusually long -- longer than with glottal stops following them.

In order to verify (a) whether legato occurred in fluent speech and if so (b) how long the legato consonant was, we inspected ten minutes of taped interviews by three different German speakers, selected from our library of foreign recordings. Both questions were answered. (a) Legato (absence of glottal stop between a final consonant and an initial vowel) was found to occur frequently. One speaker made 20 legatos out of 61 consonant-vowel junctures (41 glottal stops), another made 24 out of 69, and a third one made 18 out of 57. But those legatos occurred mostly in very common sequences such as Gibt es, Es ist, Ist es, Muss ich, Weil ich, Rat ich, Wir uns, Dich allein, etc. (b) Spectrograms of legato junctures were made. They revealed that the legato final-consonant was generally shorter than either the initial consonant of the same category or the final consonant of the same category followed by a glottal stop. We conclude, therefore, that if the glottal stop is not made, a final consonant tends to be slurred, in German as well as in English, but perhaps to a lesser extent. But normally, in German, a final consonant between vowels is distinguished from an initial consonant between vowels by its greater length as well as by the glottal stop that follows (among other factors).

We can now return to the geminate consonants and see how they are distinguished, on one hand from the initial single-consonants, on the other hand from the final single-consonants, with respect to duration.

In French, since finals and initials have equal duration, the duration cue contributes to distinguishing geminates from finals as well as from initials.

In Spanish, since finals are slightly shorter than initials, the duration cue distinguishes geminates from finals slightly better than from initials. In the case of /s/ phonemes, the distinction between geminates and finals is sharpened by the voicing and the concomitant shortening which tends to occur in finals (Las olas > [lazolas]) but not in geminates (Las solas > [lassolas]).

In English, since finals are clearly shorter than initials, the duration cue distinguishes geminates from finals considerably better than from initials.

In German, since, on the contrary, finals are markedly longer than initials, the duration cue contributes to distinguishing geminates from finals considerably less than from initials. But here the glottal stop, which is present after the finals and not after the geminates, compensates for the weakness of the duration cue, and it is fair to assume that the distinction between geminates and finals is as clear as that between geminates and initials.

OVERLAPPING

In Figs. 1, 2, 3, and 4, after each duration average, a parenthesis shows the range of duration variations on which the average is based. If the ranges of geminates are compared with those of single consonants, a few cases of overlapping are in evidence, one in Spanish, several in English and in German. Those cases were checked, and it was found that they were never produced by the same speaker; they seem to be attributable to the fact that it was not possible, in spite of efforts to that effect, to make all speakers record at the same syllabic rate or with the same degree of naturalness and the same absence of awareness of what they were recording.

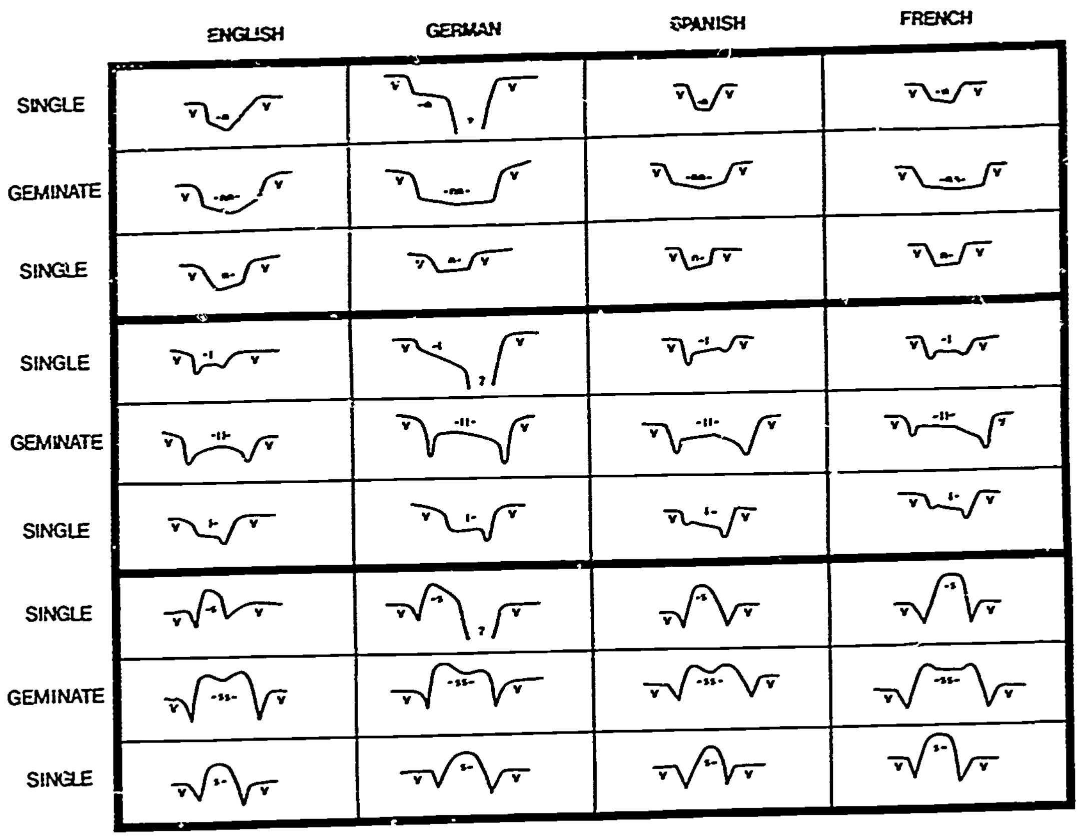
Nevertheless, there were cases in which a speaker made very little distinction in duration between a geminate and a single consonant. In three cases a speaker made a geminate only 2 cs longer than the corresponding single consonant. Yet, auditorily, the longer consonants were clearly heard as geminates. On the other hand, we came across a Spanish single /s/ of 20 cs, an English single /s/ of 21 cs, a German single /s/ of 22 cs, and an English single /l/ of 12 cs. Those are well within the range of geminate consonants in their own category (and language). Yet they are clearly heard as single consonants in the recordings.

It must be assumed, therefore, that other factors, besides duration, contribute to the opposition of geminate to single consonants. Subjective intensity or loudness is one of them.

2. THE INTENSITY FACTOR

In order to study the variations of loudness and its distribution in time at various points of the consonant, we had recorded on all our spectrograms the overall amplitude, representing the sum of the amplitudes of all harmonics at any instant. In spite of certain inconsistencies we have attempted to extract from all those amplitude lines the shapes that seem to be the most typical in each of the four languages under study. Those amplitude shapes are presented in Fig. 1 for /n/, /l/, and /s/ when these consonants are single word-final, geminate across word boundary, and single word-initial. They are in that order from top to bottom in each square of Fig. 1.

When examining these amplitude lines, one must be aware that the articulatory interpretation differs for each category of consonants. For nasals like /n/, the mouth outlet being blocked and the sound being detoured through the nostrils by way of the velic corridor, a rising line (which corresponds



AMPLITUDE DISPLAY OF SINGLE AND GEMINATE CONSONANTS

Figure 1

to an increase in loudness) must indicate an increase in air pressure from the lungs. For glides like /l/, the mouth outlet being laterally open, a rising line (increase in loudness) must indicate mainly a decrease in tongue-tip pressure against the alveols, so that the dips which can be seen at the beginning or the end of an /l/-hold must be interpreted as instants of sharp tongue pressure, and a falling line as an increase in tongue pressure. This explains why the /l/ lines rise and fall in the opposite direction of the /n/ lines. For constrictives like /s/, a rise of the amplitude line (increase of loudness) must mainly indicate an increase in air pressure from the lungs, because the central tongue-groove and the teeth slit maintain a rather fixed optimal aperture without which an /s/ cannot be produced.

As a whole, Fig. 1 shows that the variations of loudness seem to play a part in distinguishing geminate from single consonants and that they correlate with, and support, the role played by duration. Furthermore, the distribution of loudness is different in each language and contributes to characterizing each one.

The most important thing that emerges from the wealth of articulatory details in Fig. 1 is that most of the time, and perhaps always, there really are two phases in the articulation of geminate consonants, one which reflects a character of the final consonants, the other which reflects a character of the initial consonants. Thus, geminate /n/ starts with a fall and changes toward mid-course to a rise, geminate /l/ does the opposite, and geminate /s/ shows a dip toward the middle of the friction. It is especially relevant to observe also that geminate /l/ has both an arresting dip and a releasing dip. It is true that both dips are sometimes found in the Spanish and French single /l/, but only one of the two dips is sufficiently marked to compare with the dips of geminates.

In one case, Fig. 1 provides information which does not appear in the duration data of Table 4. In French, the amplitude behavior for final /n/ and final /l/ is not quite the same as for initial /n/ and initial /l/, whereas Table 4 gives similar durations to finals and initials.

For the details, Fig. 1 speaks for itself and should need no further comments.

3. THE DURATION OF THE PRECEDING VOWEL

Considering that a very close relationship exists between the duration of a voiced consonant and that of the preceding vowel -- in synthesis a voiced consonant can be made to sound voiceless simply by shortening the vowel that precedes -- we were curious to know whether geminate consonants would be preceded by shorter vowels than the corresponding single consonants. To our surprise, we found that it was not the case, or at least not in a significant manner. In English, we measured, for six speakers, the duration of /i/ before /nn/ and /n/ and of /e/ before /ss/ and /s/ with the following results (averages in centiseconds are followed by the duration of the consonant in parenthesis):

I've sEEn Nelly
10 (12.5)

I've sEEn Elly
11 (6)

We sEE Nelly
9.9 (7.5)

The rAce sends
19 (22)

The rAce ends
18.2 (12)

The rAY sends
20.1 (17)

In the other languages, vowel-duration results are comparable to those of the preceding English examples -- vowels before geminates are on the average only slightly shorter than vowels before single consonants. A rapid sampling gives the following average ratios: Spanish .94 to 1, English .96 to 1, French .96 to 1, and German .97 to 1. Divergences are not wholly inconsistent, however. They depend on the speaker, which is a random effect; but they are also related to vowel quality in a way that is perhaps not random. The [ɛ] vowels are predominantly longer before a geminate than before a single consonant in Spanish (El lápiz vs. El Apis), in French (Il acquerrait vs. Il acquérait), and in German (Nenn Nomen vs. Nenn Omen); and the same can be said of the vowel [a] in German (Sie sahn Nomen vs. Er sah Nomen) and in French (Il barrerait vs. Il barrait). Back vowels, on the other hand, are predominantly shorter when followed by a geminate in Spanish (Un nido vs. Unido) and in French (Il courrait vs. Il courait). But what is most striking as one looks at spectrograms of these utterances is the number of cases in which a vowel preserves its original length despite a practical doubling of the following consonant's duration, as in:

The rAce ends vs. The rAce sends.
17 (12) 17 (23)

We must therefore assume, to our dismay, that in distinguishing a geminate from a single consonant, the duration of the preceding vowel is a negligible factor. This suggests that consonant gemination does not involve mental anticipation of an extra effort. If it did, the preceding vowel would be sharply shortened, as it is when a voiceless consonant follows rather than a voiced one. This interpretation is based on the theory that voiceless consonants, in order to be heard without the high return help of vocal cord vibrations, require a greater expense of articulatory energy than their voiced counterparts (cf. André Malécot, "An experimental study of force of articulation," Studia Linguistica, 12: 35-44, 1958).

4. CINERADIOGRAPHY

Up to this point, our data on consonant duration was extracted from magnetic recordings made visible on acoustic spectrograms. Now we must briefly mention the articulatory aspect of gemination which can be observed on x-ray films, in normal or slow motion with sound, and frame by frame without sound.

The lists of words and utterances of Tables 1, 2, 3, and 4 are all on films taken from native speakers of each language. Looking at those films is simply fascinating, but what they reveal -- mainly the complexity of the tongue movements as they produce the sounds one hears -- is hard to bring to the reader.

Only a limited number of articulatory features can be described objectively, such as tongue contact, tongue pressure, and certain simple tongue movements.

The three consonants of our study, /n/, /l/, /s/, being apical, we have counted the number of frames which show a contact of the tongue tip with the upper incisors or the alveols. Since the films are taken at 24 frames per second, the time between two contacts of the tongue against the alveols is nearly 4 cs. Two contacts, however, cannot be interpreted in exact durations -- they may correspond to any length of time from a little more than 4 cs to a little less than 8 cs. It is interesting, nevertheless, to note their numbers and to compare them with the duration figures of Tables 1, 2, 3, and 4.

The number of conta ts (meaning the number of frames showing a tongue contact) for these three apical consonants varies (on our films) from 2 to 8. (We shall see in the second part of this study that flaps, as in the single-trill Spanish /r/, generally show only one contact.)

In English, the single /n/ of I've seen Elly and We see Nelly generally showed 2 contacts and the geminate /n/ of I've seen Nelly 4 contacts. Exceptionally, the single /n/ showed 3 contacts and the geminate /n/ 5 or 6.

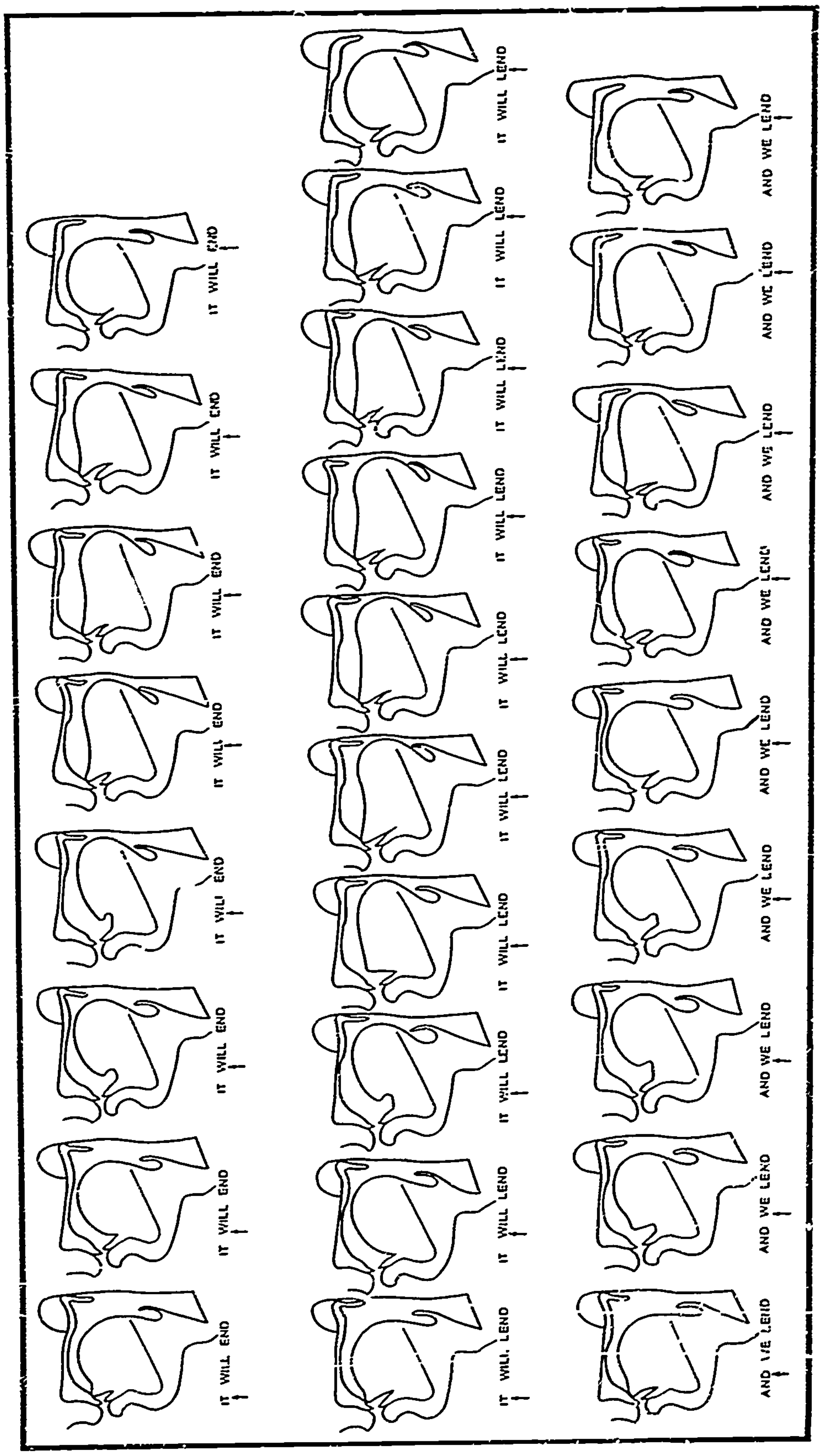
The other languages had more tongue contacts than English for /n/. They varied, like English, from 2 to 3 contacts for single consonants to 4, 5, or 6 contacts for geminate consonants, but the occurrence of 3 contacts for single and 5 contacts for geminate /n/ were much more frequent than in English. This agrees with the results of Tables 1, 2, 3, and 4 which show that in absolute time the /n/ and /nn/ sounds were shortest in English (7.9 vs. 12.2) and longest in Spanish (8.3 vs. 17.6).

Practically the same remarks and figures apply to the /l/ and /ll/ sounds. The English films show less contacts than those of the other three languages, a fact which confirms the results of Tables 1, 2, 3, and 4. This suggests that the well observed tendency for English medial /t/ and /d/ to be articulated as a flap in words like Latter and Ladder may also be at work in a slight slurring of geminate /n/ and /l/.

The /s/ sounds show more contacts than the /n/ or /l/ sounds in all the languages. Accordingly, they are much longer than /n/ or /l/ sounds on Tables 1, 2, 3, and 4. With the /s/ sound, the English film is characterized by slightly more contacts than the three other languages. In general, single /s/ is seen to make contact on 3 or 4 frames, whereas geminate /s/ makes contact on 4 to 7 frames. Obviously the slurring tendency of English apicals does not apply to voiceless sibilants.

In order to illustrate the detailed information that appears on films, Figs. 2, 3, 4, and 5 were prepared by tracing a sequence of /l/ and a sequence of /ll/ frames in each of the four languages. These tracings present a profile view of the tongue, the constrictions of which shape the resonating cavities of the mouth (vocal tract) from the lips to the vocal cords.

Fig. 2 uses a film of the three English sentences : It will end (top row), It will lend (middle row), and And we lend (bottom row). In each row the sequence goes from the last frame of /t/ or /d/ to the first frame of /ε/. In the two upper rows, consonant anticipation -- an outstanding charac-



ENGLISH SINGLE AND GEMINATE CONSONANTS

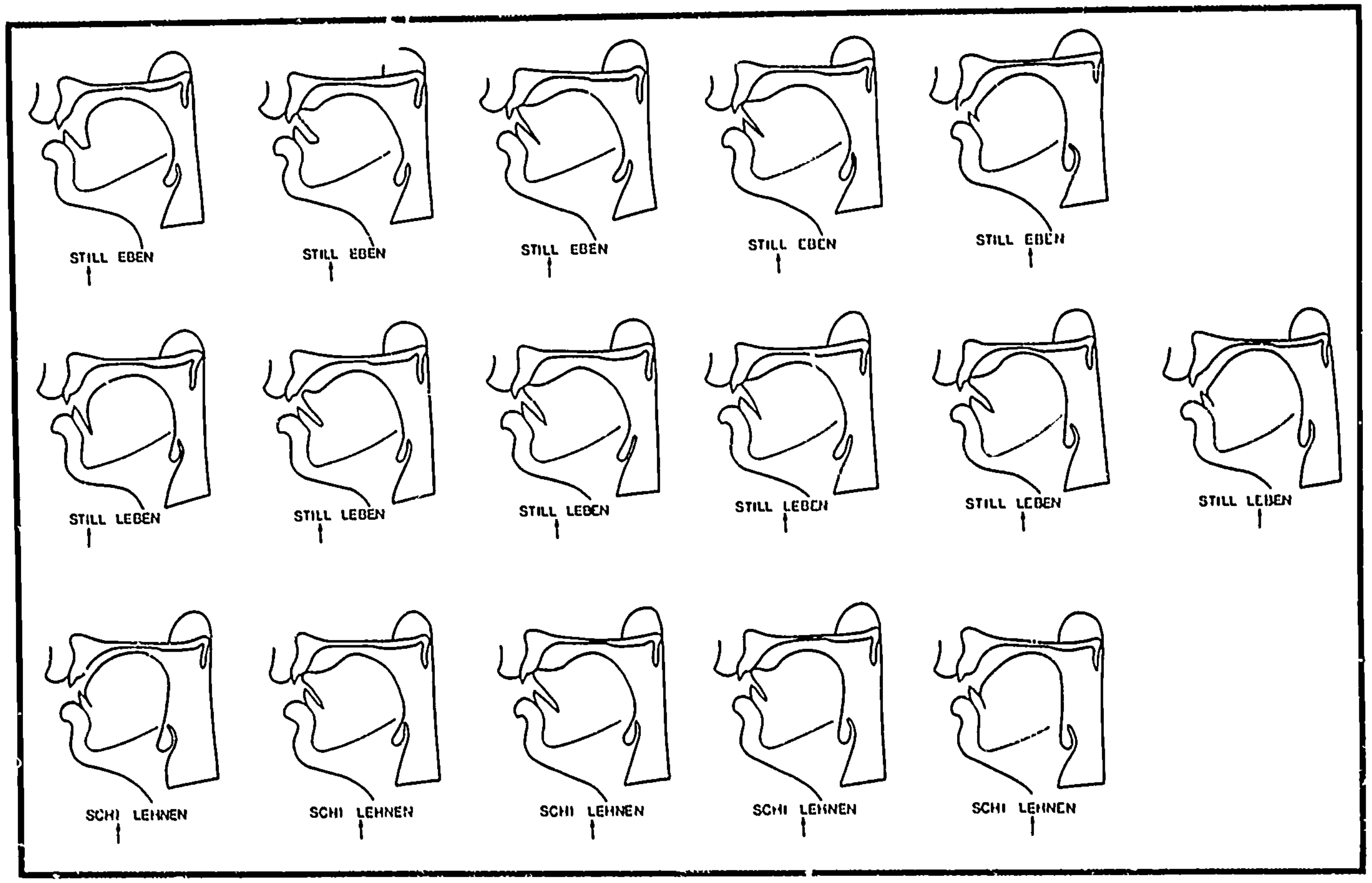
Figure 2

teristic of English phonetics -- is so strong that the /ɪ/ phoneme of Will is completely by-passed; the tongue moves directly from /w/ to /l/.

Top row: word-final /l/. The /t/ contact between the tongue tip and the alveols is abandoned rather abruptly for a /w/ position (frames 2 and 3) which requires a tongue constriction at the velum and a fairly large cavity in back of the tongue (the pharyngeal cavity) together with pronounced lip rounding (maximal at frame 3). Beginning with frame 4, while the lips are unrounding, the tongue moves toward an /l/ position in two directions at once, forward and backward -- the tip stretches forward toward the alveols while the back shifts its constriction from the velum to the pharyngeal wall. (This pronounced backing of the tongue toward the pharyngeal wall is a characteristic of English /l/, especially emphasized in the "dark /l/" or implosive /l/, but also present in the "clear /l/" of frame 7, bottom row, as compared with the clear /l/ fronting of German /l/, for instance, in Fig. 3.) At frame 7 of the top row, when the shift from final /l/ to /ε/ is about to take place, the tongue still has its weight in back and the tip is making loose contact with the alveols (as compared with frame 9, middle row). Furthermore, the separation of the tip from the alveols (frames 7 to 8) is slow as compared with the separation of frames 9 to 10, middle row.

Middle row: geminate /l/. The geminate /l/ starts nearly like the final /l/ of the top row. The tongue moves quickly to a /w/ palatal constriction, then somewhat slower to the dark-/l/ position by stretching in two directions at once -- toward the alveols in front and toward the pharyngeal wall in back. Up to frame 7, the tongue keeps the shape of a final /l/. Then, in frames 8 and 9, the tongue pressure against the alveols increases as the tongue weight shifts forward in anticipation of the following vowel as for an initial /l/. Finally the release of the tongue tip is fast as shown by the wide difference between frames 9 and 10. Naturally, the number of tongue-tip contacts with the alveols is larger than in the top and bottom rows -- 5 contacts vs. 3, but more interesting is the double articulation of this geminate /l/ -- it clearly includes an initial /l/ phase and a final /l/ phase.

Bottom row: initial /l/. Between /w/ and /l/, here, the tongue position for /i/ is not by-passed, as it is for a final or a geminate /l/ in the two upper rows. The /w/ position is reached at frame 3 -- the lips are rounded and the tongue makes a constriction at the velum while maintaining a large pharyngeal cavity. Then the tongue moves forward for /i/, the closest to an /i/ shape being reached at frame 5. In the next moves the tongue will stretch both ways, but it reaches forward (frame 6) before reaching backward (frame 7), contrary to what occurs in the two rows above where the tongue reaches backward first. Finally (frames 8 and 9), the weight of the tongue shifts forward and the tip exerts pressure against the alveols in preparation for a release that seems sharp (frames 8 to 9) but is not so sharp as for the geminate /l/ at the end of the middle row.



GERMAN SINGLE AND GEMINATE CONSONANTS

Figure 3

Fig. 3 uses a film of the three German utterances: Still eben (final /l/), Still-Leben (geminate /l/), and Schi Lehnen (initial /l/). In each row, the sequence goes from the last frame of /ɪ/ or /i/ to the first frame of /e/.

Let us first make some general observations. Here the tongue is much more fronted than in English -- the root of the tongue remains quite distant from the pharyngeal wall. The German /l/ shape is characterized by a tongue-tip contact against the alveols and a slight bulge of the tongue root toward the back wall of the pharynx.

Top row: final /l/. A comparison of the top and middle rows presents an interesting problem. Final /l/ shows as many tongue contacts (4 frames) as the geminate /l/ of the middle row, yet it is clearly distinct from the geminate /l/, auditorily. The meaning Still eben is clear not only, perhaps, because of the glottal stop that is heard soon after the /e/ position has been assumed in frame 6 of the top row, but also because the tongue remains in final /l/ shape, maintaining its radical bulge toward the pharynx, and loose tongue contact until the tongue tip separates from the alveols. In other words, the tongue does not show any sign of /e/ anticipation as will be the case for the initial /l/ of the bottom row.

Middle row: geminate /l/. In the middle row, the tongue begins as for final /l/, but in frame 5 it shifts to a shape for initial /l/ by increasing its tongue-tip pressure, shifting its weight toward the front and abandoning the radical tongue bulge in anticipation of the /e/.

Bottom row: initial /l/. Similar anticipation of the

/e/ appears in the bottom row for the initial /l/ when the tongue root is fronted, at frame 4, and tongue-tip pressure prepares a sharp release (frames 4 to 5).

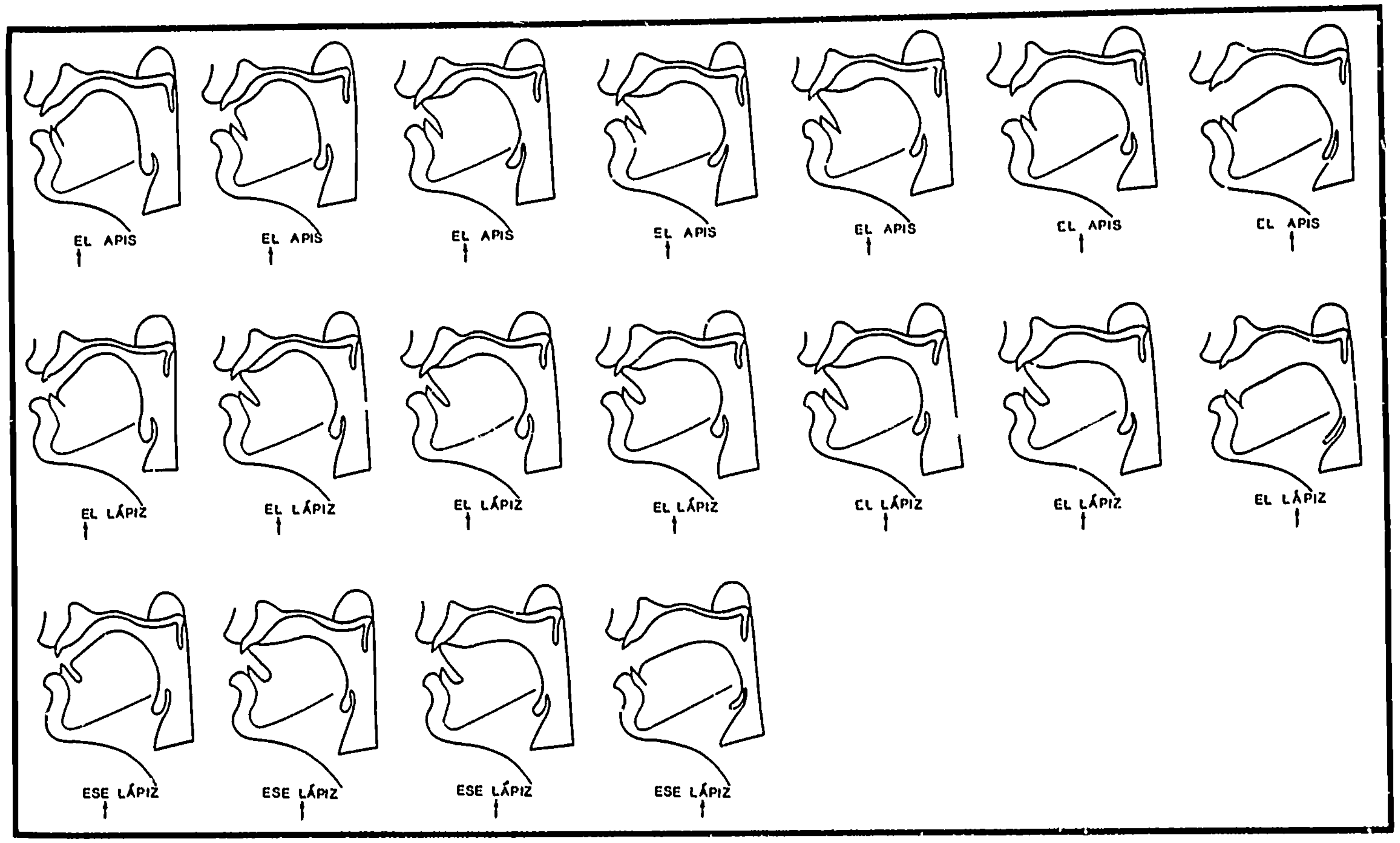
Thus in the middle row, a geminate /l/ is perceived because two different phases of articulation are produced in rapid sequence -- the final /l/ phase and the initial /l/ phase.

Fig. 4 uses a film of three Spanish utterances : El Apis (final /l/), El lápiz (geminate /l/), and Ese lápiz (initial /l/). In each row, the sequence goes from the last frame of /e/ to the first or second frame of /a/.

The fronting of the tongue is about similar to that of German. Here the number of frames with tip contact differs from row to row -- the geminate /l/ has five contacts, the final /l/ has three, and the initial /l/ two.

Top row: final /l/. The fronting of the tongue tip and the slight bulging of the tongue root to assume the /l/ position occur in frames 2 and 3. Once that position has been reached, it is kept to the end without further bulging of the tongue root until the tongue tip separates from the alveols. Since the /a/ that follows requires a marked bulge of the tongue root, the lack of bulge in frame 5 clearly indicates a lack of anticipation of the following vowel.

Bottom row: initial /l/. The characteristic tongue position is reached at frame 2, and in frame 3 the radical bulge of the tongue toward the pharyngeal wall is slightly extended, while the dorsum is lowered, in anticipation of the following /a/ which is about to acquire a pronounced pharyngeal constriction.



SPANISH SINGLE AND GEMINATE CONSONANTS

Figure 4

Middle row: geminate /l/. The geminate /l/ includes both the final and the initial phase. The tongue position of /l/ is assumed first by raising the tip (frame 2), then by lowering the dorsum (frame 3); and the /a/ position is anticipated in frames 5 and 6 by a continuous backing of the radical bulge and lowering of the dorsum.

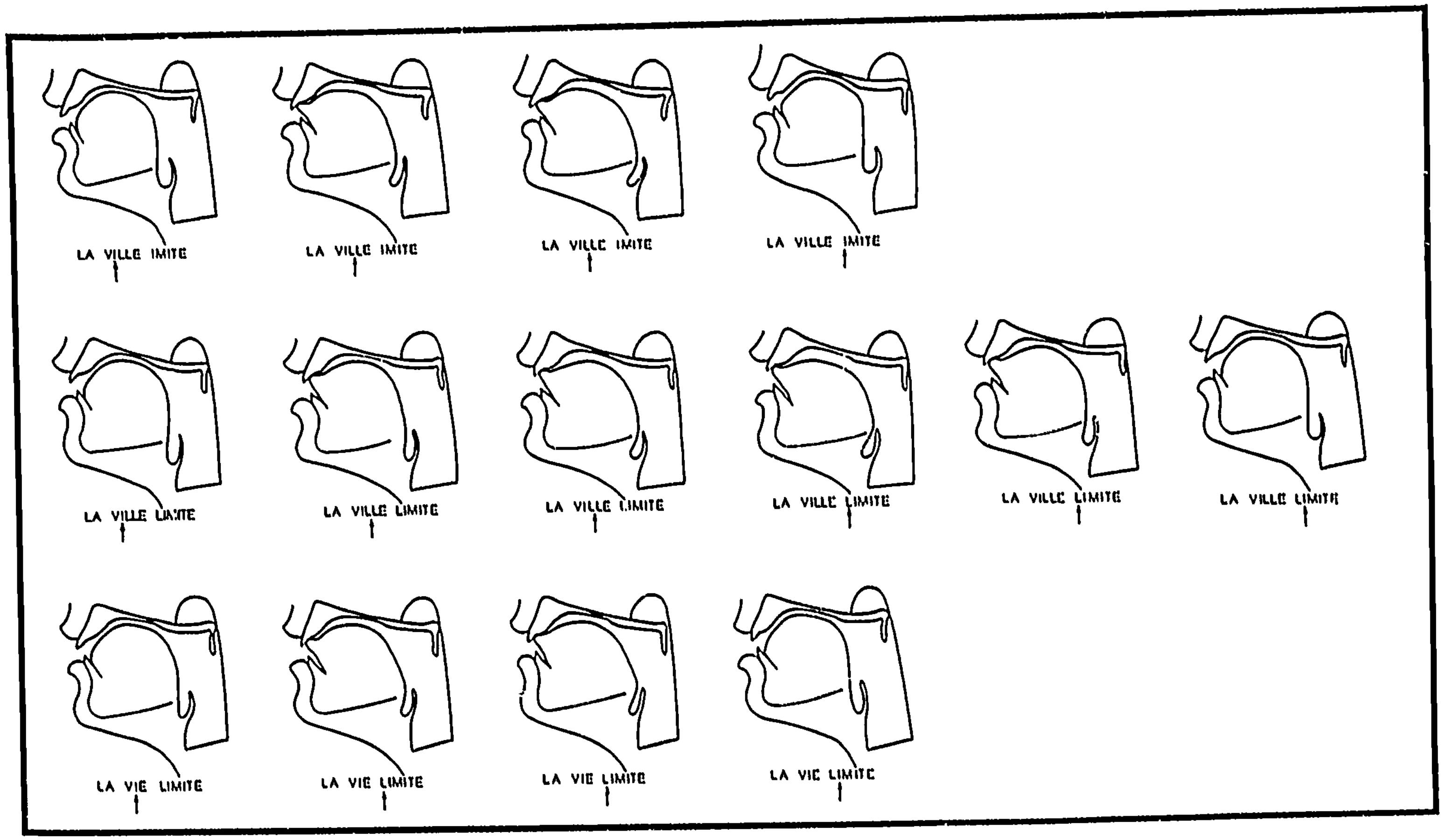
Fig. 5 uses a film of three French utterances: La ville imite (final /l/), La ville limite (geminate /l/), and La vie limite (initial /l/). In each row the sequence goes from the last frame of the first /i/ to the first frame of the second /i/.

Here, the characteristic position of the tongue for /l/ is very divergent from that of English. The tongue is even more fronted than for German or Spanish, the tip is clearly dental rather than alveolar, and the dome of the tongue is maximally high and bulging in contrast to the sagging tongue-dorsum of the English /l/.

Top row: final /l/. The tongue tip reaches the /l/ position in frame 2, but the radical tongue-bulge reaches it only in frame 3. A lack of anticipation of the vowel /i/ which follows is shown by the radical bulge being maintained as long as the tip has not left the alveols.

Bottom row: initial /l/. The characteristic /l/ position is already assumed at frame 2 with the tip contact as well as the radical bulge. But at frame 3 the bulge disappears, the tongue being fronted in anticipation of the following vowel /i/.

Middle row: geminate /l/. Both the final and the initial phase appear here -- the tongue assumes the /l/ shape gradually in frames 2 and 3, then shows anticipation of the vowel /i/ by



FRENCH SINGLE AND GEMINATE CONSONANTS
Figure 5

fronting in frame 5 before the tip has left the alveols.

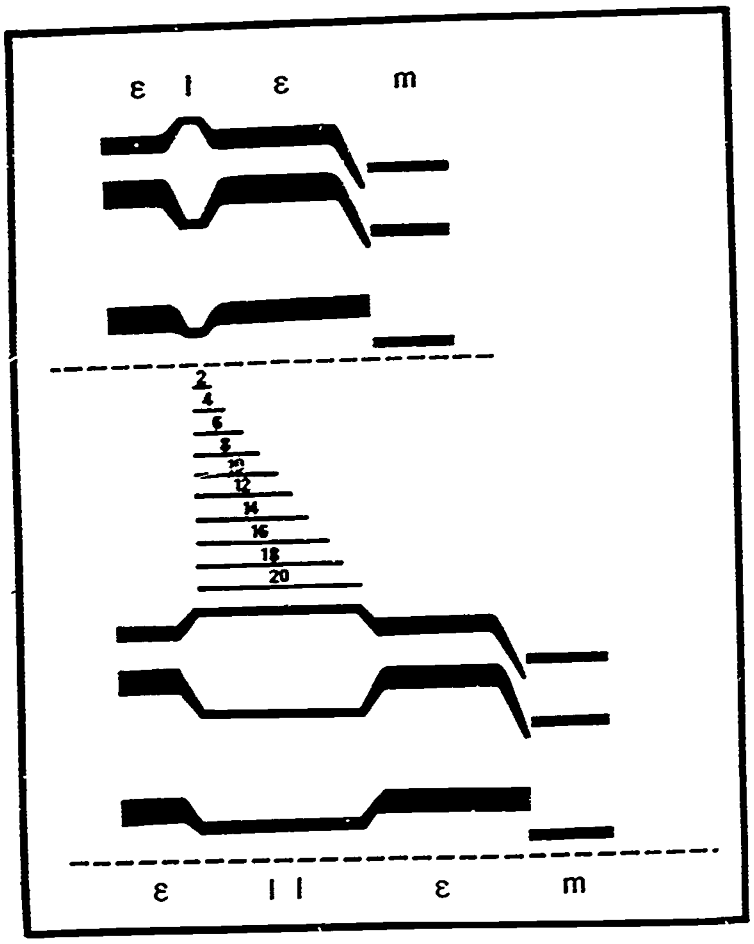
Thus, as was the case with the amplitude recordings of Fig. 1, the x-ray frames reveal certain subtle differences of articulation between initial and final /l/ in French, which do not appear in the duration data of Table 4.

5. THE PERCEPTUAL TEST

The acoustic and articulatory analysis of gemination can be complemented by perceptual tests of controlled variables produced by synthesis. The research technique of speech synthesis makes it possible to isolate one of the acoustic correlates of speech in order to vary it separately from the others and to judge by ear the effects of all changes. Here we chose to test first the factor of consonant duration in distinguishing geminate from single consonants. Let us call this Test A. Using an artificial-speech machine, we synthesized the utterances: Elle aime, I' n'ira pas, Laisse Elie, which can be heard as Elle l'aime, I' n' niera pas, Laisse ces lits when the consonant hold is sufficiently prolonged. We made ten versions of each synthetic pattern, giving to each version a different consonant duration in such a way as to cover the range of hold durations which includes single and geminate consonants of the appropriate category. For /l/ and /n/ the length of the consonant-hold was given durations varying from 2 to 20 cs in steps of 2 cs. For /s/, the range went from 8 to 26 cs in steps of 2 cs, 8 cs being the low limit for voiceless fricatives (for 6 cs and below, fricatives are mostly heard as voiced in the four languages under study). In order to make the comparison among languages more valid, we used the same words for all tests. These words were French because it was possible to find, locally, native speakers of the three other languages who understood enough French to distinguish between simple pairs of utterances, yet who had not lost their native habits of speaking and hearing.

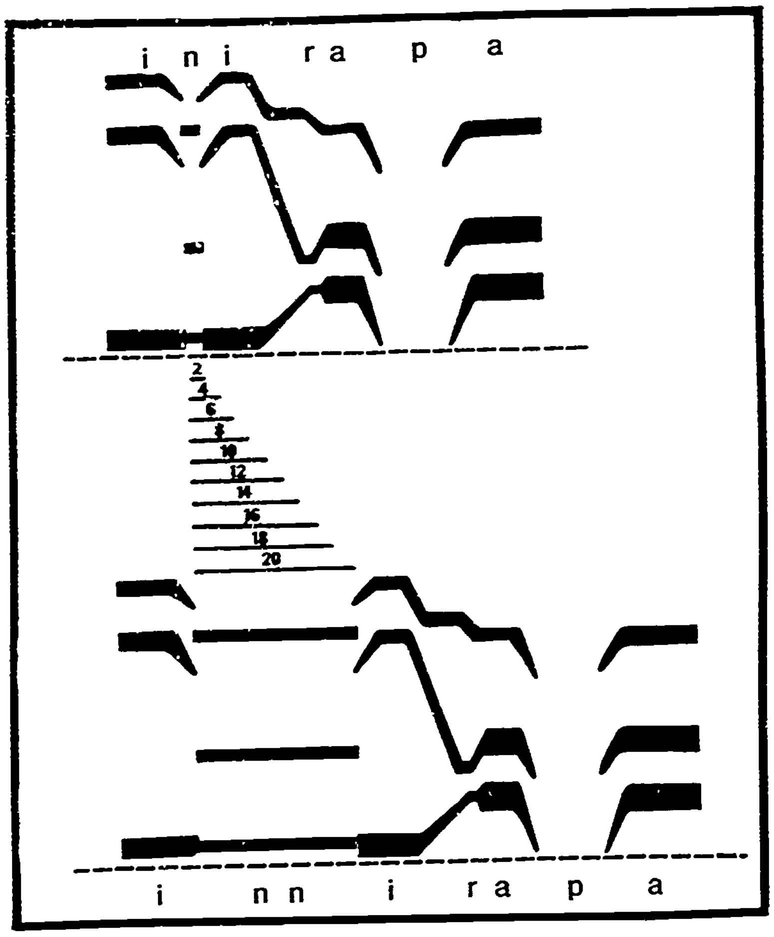
Figs. 6a, 6b, 6c present the patterns in their shortest and longest versions with lines indicating the consonant length of the eight intermediary patterns. Each of the ten patterns was transformed into sound and recorded five times. The 50 versions of each sentence were mixed in random order to make three separate tests, and the stimuli of each test were presented for judgement by ear to three native speakers of English, German, Spanish, and French. In each test they were given a sheet of paper on which were printed 50 pairs of utterances such as Laisse Elie, Laisse ces lits, and they were asked to circle the one they had understood for each stimulus. The stimuli were heard only once and came in quick succession in order to force rapid judgements without reflexion or hesitation.

The results of these tests showed the same sort of divergences among languages as the measurements of Tables 1, 2, 3, and 4. No crossover points appeared clearly between geminate durations and single-consonant durations. Rather, the geminate were separated from the single consonants by a wide range of ambiguous durations -- durations that were nearly all heard indifferently as single or geminate. For /n/ and /l/ heard by American listeners, the range of ambiguity was the lowest, that is, 6, 8, and 10 cs. Below that range, all /n/ and /l/ consonants were perceived as single, and above that range all /n/ and /l/



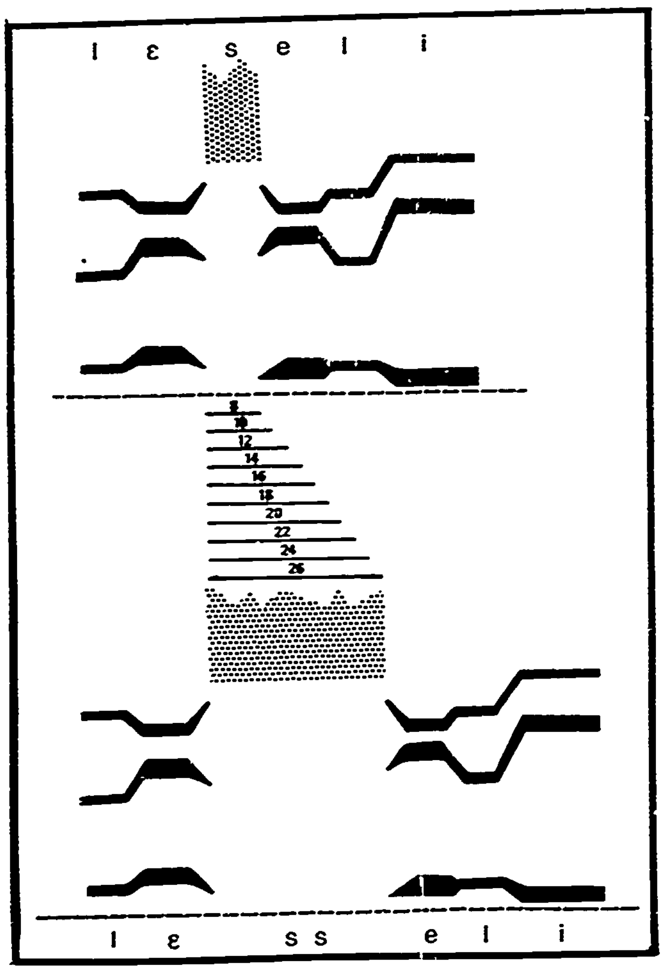
TIME VARIATIONS IN SYNTHETIC PATTERNS

Figure 6a



TIME VARIATIONS IN SYNTHETIC PATTERNS

Figure 6b



TIME VARIATIONS IN SYNTHETIC PATTERNS

Figure 6c

consonants were perceived as geminate. For the same consonants heard by German, Spanish, and French listeners, the range of ambiguity was 8, 10, and 12 cs.

For /s/ heard by American listeners, the range of ambiguity was the highest, that is, 16, 18, and 20 cs. For /s/ heard by French listeners, the range of ambiguity was 14, 16, and 18 cs. And for /s/ heard by Spanish and German listeners, the range of ambiguity was 10, 12, 14, and 16 cs. This low range may be due to problems of voicing in those languages -- initial /s/ always being voiced in German and final /s/ often being voiced in Spanish, geminates ought to be shorter than in languages where both elements of the geminate are clearly voiceless.

Secondly, an exploratory test was made of the contribution of formant transitions in recognizing gemination. Let us call it Test B.

The same sentences were used as in Test A, but with /n/, /l/, and /s/ modified in such a way that either the arresting formant-transitions or the releasing formant-transitions suffered reduction of intensity of about 10 decibels at any instant.

The results of Test B were not strikingly different from those of Test A. Stimuli in which /n/, /l/, and /s/ had been heard unambiguously as geminate in Test A were still heard as geminate in Test B, showing, perhaps, that the duration factor tends to override the transition factor. But among the stimuli that had been heard ambiguously, 15 per cent more than in Test A were judged as single consonants.

More interesting was the result that, for American subjects, the reduction of the releasing transitions had clearly more effect than the reduction of the arresting transitions; whereas for German, Spanish, and French subjects it was the reverse -- the reduction of the arresting transitions had more effect than the reduction of the releasing transitions. This is, perhaps, an indication that in English the initial-consonant phase of articulation contributes more than the final-consonant phase in distinguishing geminate from single consonants, and that the reverse occurs in the other languages. Naturally, the fact that final-consonant anticipation is more marked in English than in the other 3 languages seems to be related to our new finding. This relation suggests that speakers of a given language preferably recognize gemination by the addition of the consonant phase that is the least common in that language. This theory would, of course, need verification with other languages.

B. GEMINATION WITHIN WORD BOUNDARY

Spanish, French, and German each have their particular problems with respect to gemination (or lengthening) of the consonant /r/ within word boundary. It is noteworthy that /r/ is the only consonant capable of geminating or lengthening meaningfully within word boundary, in Spanish, French, and German, and that this capability is not found in English.

FRENCH r/rr

The problem of gemination is simplest in French. There, three verbs and their components clearly distinguish between the imperfect indicative and the conditional present by gemination of the medial /r/ of three verbs: mourir, courir, and acquérir, [il murɛ], [il kurɛ], [il akɛrɛ] meaning He was dying, He was running, and He was acquiring, whereas [il murrɛ], [il kurrɛ], [il akɛrrɛ] mean He would die, He would run, and He would acquire. In addition, all verbs in which final -rer occurs after a vowel or another -r, such as désirer, honorer, barrer, and serrer, normally drop the [ə] between /r/'s and present the same type of oppositions, Il désirait [dezirɛ] meaning He was wishing, whereas Il désirerait [il dezirrɛ] means He would wish. There are nearly 300 such verbs.

For French, we made motion picture x-rays and tape recordings of seven minimal pairs as spoken by three native speakers of French. The pairs are listed below with phonemic transcriptions.

Il mourait /mure/	Il mourrait /murre/
Il courait /kure/	Il courrait /kurre/
Il acquérait /akere/	Il acquerrait /akerre/
Il désirait /dezire/	Il désirerait /dezirre/
Il honorait /onore/	Il honorerait /onorre/
Il serrait /sere/	Il serrerait /serre/
Il barrait /bare/	Il barrerait /barre/

DURATION VARIATIONS

Durations of the French geminate and single /r/ sounds were measured in centiseconds on spectrograms and averaged. The results are consistent and quite similar for the three speakers. The geminate consonants averaged 20.4 cs and the single consonants 11.2 cs, for a ratio of 1.8 to 1 which is comparable to the ratio of 1.9 to 1 found for the French /n/, /l/, and /s/ combined. The absolute durations (20.4 vs. 11.2) place the /r/ sounds higher than the nasals (15.5 vs. 8.2) or laterals (16.8 vs. 7.5) but lower than the voiceless constrictives (21 vs. 13.4) on a scale of duration. According to this, the /r/ sounds are the longest of the resonants in French.

A very interesting side result emerges here. The geminate /r/ divides into two groups according to spelling. Geminates that are spelled -rer, as in désirerait or barrerait, were regularly longer, for all speakers, than geminates that are spelled -rr, as in courrait. The courrait type averaged 18.8 cs, whereas the désirerait type averaged 22.0. This suggests, of course, that the unstable [ə], which is supposed to "fall" when preceded by a single consonant, still shows signs of life between /r/ sounds.

INTENSITY VARIATIONS

The amplitude line on the spectrograms leaves no doubt that the French /rr/ consonants of our study are geminates and not long consonants. Whereas the single /r/ shows a single dip of amplitude (in fact a deep groove), the geminate /r/ always shows two dips in the amplitude line, one at the beginning (an arresting depression) and another at the end (a releasing depression). The releasing depression is deeper than the arresting one and often corresponds to an instant of unvoicing in the spectrogram's formants. Between the two depressions a dome appears whose general trend is more often falling than rising, an indication that the releasing phase is slightly more stressed than the arresting one.

Signs of uvular flaps are at times visible on the amplitude line of the geminate /r/, never on the amplitude line of the single /r/. But they are too occasional in appearance, too irregular in time intervals to be counted as a contribution to the geminate vs. single distinction. They do not have either the clarity or the periodicity (equal intervals of time) of the Spanish apical /r/ or of the German uvular /r/, but appear rather as an irregular variation in the noise disturbance caused by the proximity of the tongue root to the uvula.

SPANISH r/rr

In Spanish, as is well known, multiple flap /r/ occurs medially, where it contrasts with single-flap /r/ (Carro vs. Caro), as well as initially (Raro), where it does not contrast but is conditioned by position (single-flap /r/ does not normally occur in word-initial position). The multiple flap /r/, therefore, is distinctive in one position and not in the other -- a problem of interest to the phonetician.

For Spanish, we made motion picture x-rays and tape recordings of eight minimal pairs of the medial contrast: multiple-flap /r/ vs. single-flap /r/, and eight words in which multiple-flap /r/ appears initially before the same vowels as the medial /r/'s. The list of words follows:

Carro	Caro	Rama
Barra	Bara	Raza
Perro	Pero	Remo
Cerro	Cero	Reza
Corro	Coro	Roma
Forro	Foro	Roza
Querría	Quería	Rico
Arrugas	Arugas	Ruga

Three native speakers of Spanish were used for the recordings, one from Mexico and the two others from Spain. With respect to /r/, their pronunciation was rather uniform.

DURATION VARIATION

Both the duration and the number of flaps were measured for all /r/ sounds. Single-flap /r/'s average 4.3 cs in our recordings. Multiple-flap /r/'s in medial position average 13.5 cs and 3.8 flaps. Multiple-flap /r/'s in initial position average 15.1 cs and 3.3 flaps. Spectrograms show clearly why initial /r/'s are longer in spite of having less flaps than geminate medial /r/'s. In initial position the first interruption is preceded by a vocalic period of preparation during which the intensity rises to a level just high enough for the first dip not to hit bottom. On the x-rays this period of preparation corresponds to about one frame during which the tongue tip stands raised awaiting the air-flow that will make it vibrate.

Our duration data are in good agreement with earlier studies by Navarro Tomás, made by means of the kimograph in Madrid

(*Revista de Filología Española*, V, 387).

INTENSITY VARIATIONS

The amplitude line shows a clear dip (groove, depression) for every apical flap (partial interruptions of the air-flow by periodic contacts of the tongue tip with the alveols).

Naturally, a single flap produces a single dip. With the occurrence of multiple flaps, the dips are of equal depth and appear at equal intervals.

Gemination (rather than prolongation) is not as clearly indicated as in French. Not always, but in more than half of the occurrences of intervocalic /rr/, the first and, to a greater extent, the last dip are wider and deeper than the others, an indication of two separate phases, one arresting, the other releasing. But between the first and last dip, the overall shape of the amplitude is not that of a dome, as in French, but rather that of a straight line which tends to rise when a stressed vowel follows, as in *Torrija*, and to fall when an unstressed one follows, as in *Forro*.

Multiple flaps, in word-initial position, offer a totally different picture. The dips regularly occur along a sharply rising overall line which goes from a zero level to the amplitude level of the first-syllable vowel. The dips are equal; therefore, no indication of gemination appears, and one is perhaps justified in calling the word-initial /r/ of Spanish a long /r/ rather than a geminate /r/.

GERMAN r/rr

In German, the problem is, perhaps, not one of gemination but of force of articulation or prolongation. After short vowels, either in final or in medial position, the German /r/ is generally stronger and longer than after long vowels. This contrast is quite pronounced in final position because of the remarkable weakness of final /r/ after long vowels, but it can also be clear in medial position.

To study this we have made motion picture x-rays and tape recordings of the following minimal (or near minimal) pairs:

Star	starr
Heer	Herr
wir	wirr
ihr	irr
schmoren	schnorren
Behaarung	Beharrung

Obviously, the differences in vowel length and/or in vowel quality make the differences in /r/ articulation appear redundant. But one might argue that it is the difference of vowel that is redundant because, historically, it is the strong consonant that caused the vowel to shorten, and later the shortening that caused a change in vowel quality. Whatever the conjecture, it is evident that the /r/ differences contribute to distinguishing the words on the left from the words on the right.

Eleven German speakers were asked to record these minimal pairs in random order and with other words mixed in. Seven of those speakers used the uvular /r/ and four used the apical /r/.

DURATION VARIATIONS

Duration measurements were made on spectrograms not only for the consonants, but also for the vowels that precede them. In the results given below, the length of the preceding vowel is given in parenthesis. All numbers are in centiseconds and represent averages of all the words by all the subjects. Differences were very consistent.

Uvular /r/

	Ratios
Final weak /r/ (after long vowel of 19.8): 13.3	1.8 to 1
Final strong /rr/ (after short vowel of 10.8): 23.8	
Medial weak /r/ (after long vowel of 19.3): 10.9	1.4 to 1
Medial Strong /rr/ (after short vowel of 8.4): 15.0	

Apical /r/

Final weak /r/ (after long vowel of 23.0): 12.6	1.6 to 1
Final strong /rr/ (after short vowel of 11.2): 20.9	
Medial weak /r/ (after long vowel of 18.3): 8.2	1.5 to 1
Medial strong /rr/ (after short vowel of 9.1): 12.5	

In final position, the duration ratio of strong to weak /r/ is more significant (1.8 to 1 and 1.6 to 1) than for geminate /n/, /l/, and /s/ across word boundary (1.5 to 1 in Table 2). In medial position, the duration ratio of strong to weak /r/ is slightly less significant (1.4 to 1 and 1.5 to 1) than for /n/, /l/, and /s/ across word boundary (1.5 to 1 in Table 2). We must conclude that the duration difference between these strong and weak /r/'s is sufficiently pronounced to make the distinction perceptually possible without the help of differences of length or color in the preceding vowel.

INTENSITY VARIATIONS

In medial position, the difference between weak and strong /r/ (Beharrung vs. Behaarung) is only a matter of length -- for weak /r/, the amplitude depression is shorter, for strong /r/, it is longer. Evidence of uvular trills may appear in either case, but there are no signs of gemination in the strong /r/; it is only longer.

In final position, the amplitude line of weak /r/ is very different from that of strong /r/. For the weak /r/, while the formant transitions show that vowel quality shifts from the final-vowel quality toward a near-/a/ quality, the amplitude line remains high (after a low-intensity vowel, as the /i/ of wir, it even rises), then it falls fast to zero at a 45 degree angle. In the strong /r/, the high amplitude of the vowel is followed by a sharp dip to a mid amplitude, a plateau showing uvular trills follows, and finally a sharp drop to zero occurs.

The difference between the two /r/'s is so pronounced that when the film is played in reverse, wirr and Herr are heard as [rrɪv], [rrɛç], whereas wir, Heer return a very diphthongal [xaiv], [xaeç]. The formant spectrum shows that, for the weak final /r/, the formants shift to a vowel very close to [a]. In Heer, for instance, the first formant rises as the second falls until an /a/ position is reached, and the two formants retain that [a] position until, at the very end, they unvoice. In Herr, on the contrary, there is no time given to an [a] vowel before the constriction; uvular beats start immediately.

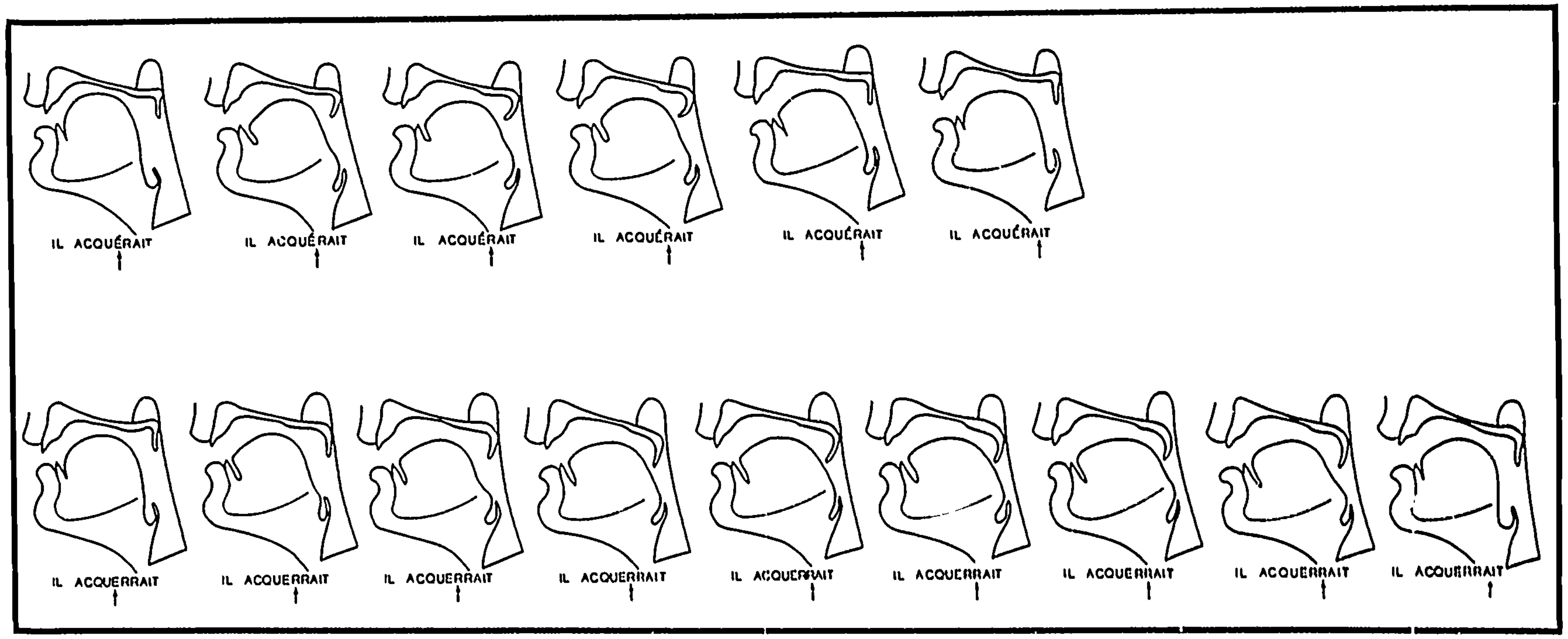
However, there is no evidence of gemination in the strong final /r/; it is only longer, more noisy, more interrupted, has different formants, different formant-transitions, and a different distribution of intensity than the 'weak' /r/. No single objective word can combine all these factors. 'Stronger, less vocalic' are only subjective notions.

CINERADIOGRAPHY

To complement and make more concrete our description of the r/rr oppositions, we have made tracings of x-rays for one speaker of each language. They are in Figs. 7, 8, and 9.

In Fig. 7, the French /r/'s are contrasted in the words Acquérait (single) and Acquerrait (geminate). Each of the two sequences starts at the last frame of [e] and ends at the first frame of [ɛ]. The /r/ of this subject is characterized by a bulging of the tongue root, near the epiglottis, toward the back wall of the pharynx, as can best be seen in frames 5 of the upper row or 7 of the lower row. A tongue bulge somewhere along the wall of the pharynx (often higher than here) must appear and divide the mouth into two cavities if the /r/ sound is to be produced. A curling up of the uvula also occurs to make trilling possible, but this is not indispensable for the production of French /r/.

The French single /r/, here, shows only the releasing phase of an initial consonant: the tongue moves quickly to the /r/ position (frames 1 to 2), maintains it for three frames (2, 3,



FRENCH SINGLE AND GEMINATE r

Figure 7

and 4), then moves forward to anticipate the next vowel.

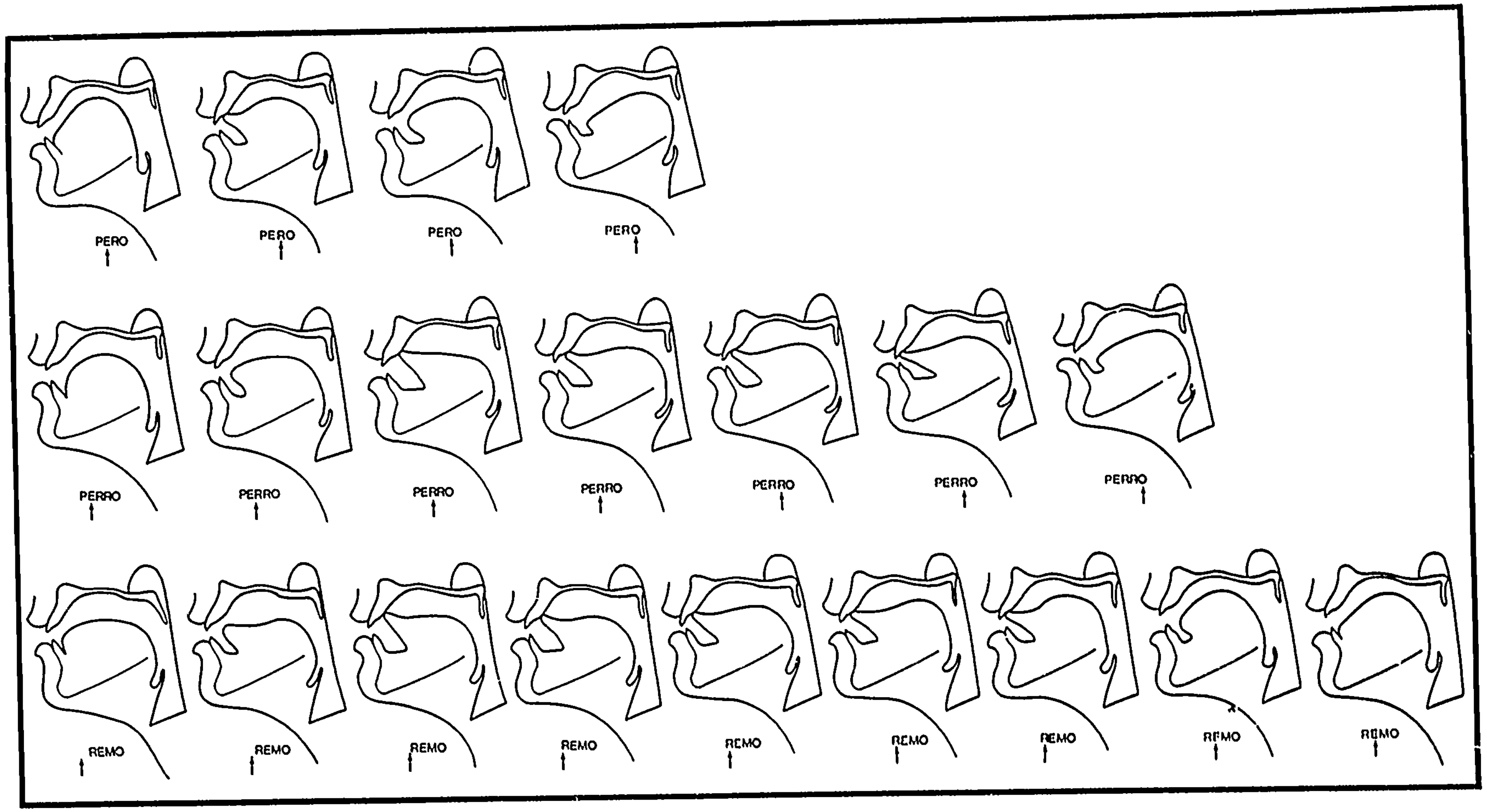
The French geminate /r/ (lower row) shows two phases. In the arresting phase (final consonant), the tongue moves slowly to the /r/ position (frames 1 to 4); in the releasing phase, which ends at frame 8, the high tongue-dorsum moves forward in anticipation of the [ε] vowel that follows.

Another difference is in the greater narrowness of the constriction between the root of the tongue and the pharynx for the geminate consonant. The narrowness of the pharyngeal constriction, just before release (frame 8), is reflected on spectrograms as a minimum of intensity.

Fig. 8 presents the Spanish /r/'s of Pero (single flap, medial), Perro (multiple flap, medial), and Remo (multiple flap, initial). For Pero, between the last frame of /e/ and the first frame of /o/ the tongue tip contacts the alveols in only one frame, and this contact is fronted like the last ones of Perro or Remo. The single-flap is, therefore, more like the releasing phase of a geminate than like its arresting phase.

For Perro, two phases are visible. In the arresting phase (frames 1-3) the tongue tip rises slowly and loosely to contact the alveols; and in the releasing phase the back of the tongue withdraws strongly toward the pharynx in anticipation of the /o/. We note also that in Perro the contact of the tongue tip begins high behind the alveols (frame 3) and moves gradually forward toward the teeth (frames 4, 5, and 6) as the flaps are being produced.

For Remo, since nothing precedes, the mouth is shown first in breathing position with the tongue at rest. In the second



SPANISH SINGLE AND GEMINATE r

Figure 8

frame the velum is half-way toward closing and the dorsum of the tongue is flattening in preparation for the concave shape it assumes in frame 3. At frame 4 no flaps occur -- the tongue is awaiting the flow of air that will set its tip in elastic motion. The flaps begin, at the earliest at frame 5 and continue until frame 8, when the dome of the tongue rises in anticipation of the /e/ vowel (frames 9 and 10). Note that from frames 4 to 8, the tongue tip gradually lowers its place of contact as it does for Perro. This forwarding of the tongue tip during the production of a multiple-flap /r/ is found in all our speakers.

Fig. 9 contrasts the weak final German /r/ of Heer after a long vowel and the strong final German /r/ of Herr after a short vowel.

For the weak /r/ of Heer (top row) the tongue changes slowly from an /e/ palatal constriction to an /a/ low-pharyngeal constriction (frames 1 to 3). At frame 3, the tongue shape is quite that of an /a/, but the narrow jaw-aperture makes that /a/ slightly obscure. At frame 4, the pharyngeal constriction has risen to about mid-pharynx. At frame 5, it has risen to the upper pharynx and the velum has lowered and placed itself along the tongue to produce a little fading-friction which, when played in reverse, makes the word Heer sound [xaeç]. At frame 6, the velum begins its breathing type of opening.

For the strong /r/ of Herr, the tongue moves directly but slowly from the /ɛ/ position to the /r/ position (frames 1 to 4) reaching a high pharyngeal constriction without having first produced the low pharyngeal constriction of Heer. During frames

GERMAN SINGLE AND GEMINATE r

Figure 9

5 to 9, the uvula approaches the tongue, making strong trills possible (uvular interruptions of the /r/ formants). Note that the uvula changes from a flat shape to a curled one, constantly following the tongue back as it moves forward. The uvula can produce trills in the flat vertical position as well as in the curled up one. (Some speakers of German use only the flat approach to the tongue, others only the curled one.) At frame 10, the velum is beginning to assume the breathing position.

In brief, the r/rr opposition is realized quite differently in each language.

In Spanish, both /r/'s are apical; but the multiple-flap /rr/ has more than three times the duration, more than three times the number of flaps of the single-flap /r/, and is articulated in two phases rather than one -- a final-consonant and an initial-consonant phase -- which can be observed not only in the amplitude variations but in the fronting of the tongue tip along the alveols.

In German, the strong /rr/ and the weak /r/ are both articulated in the pharynx, yet they seem to be different in nature. Besides being longer, the strong /rr/ shows turbulence (reflecting a narrow constriction), rapid frequency shifts of the formants (reflecting an abrupt backing of the tongue toward the high pharynx), and periodic interruptions of the formants (reflecting elastic uvular trills). The weak /r/ shows no turbulence, no rapid shifts, no interruptions, but a slow , diphthongal formant-shift to a more open vowel (reflecting a withdrawal of the tongue root toward the low pharynx), followed by a reduction of intensity which occurs when, at the very end, the low pharyngeal constriction rises along the pharyngeal wall to approach the uvula or the velum in the upper pharynx.

In French, the difference is, perhaps, less marked than in Spanish or German. Both /rr/ and /r/ show a constriction between the tongue root and the pharyngeal wall. But the geminate /rr/ is longer, more noisy (reflecting a narrower constriction), and includes two phases (a final-consonant phase and an initial-consonant phase) rather than one. In short, the distinction involves differences of quantity rather than differences of quality.

SUMMARY

Consonant gemination in English, German, Spanish, and French is investigated, (A) across word boundary, as in English We lend or We'll end vs. We'll lend; (B) within word boundary, as in Spanish Caro vs. Carro.

(A) In order to determine the acoustic, articulatory, and auditory correlates for the perception of consonant gemination across word boundary, geminated /n/, /l/, /s/ are compared with final and initial /n/, /l/, /s/ in contrasting pairs such as, German Still-Leben vs. Still eben, Spanish Es el lápiz vs. Ese lápiz, French, La ville limite vs. La ville imite, (a) by measuring the duration of these consonants on spectrograms; (b) by measuring the duration of adjacent vowels on spectrograms; (c) by analyzing the shape of intensity variations on amplitude displays; (d) by observing, frame by frame, on x-ray motion pictures, the movements of the tongue; (e) by varying the length of the consonant hold in controlled

artificial-speech patterns; (f) by varying the intensity of arresting and releasing formant-transitions.

(a) Consonant duration is found to be a major attribute of gemination across word boundary in all four languages, but the duration contrasts are wider in the two Latin languages than in the two Germanic ones, and are narrowest of all in English. Ratios between geminate and single consonants vary between 1.9 to 1 for French and 1.4 to 1 for English. (b) The duration of the preceding vowel is not a factor in the perception of consonant gemination. Vowels are not significantly shorter before a geminate than before a single consonant. (This is unexpected because vowels are shorter before a voiceless consonant than before a voiced one -- an analogical condition with respect to the anticipation of a great effort.) (c) Variations of intensity play a definite role in distinguishing geminates from single consonants. They show two phases in the geminated /n/, /l/, /s/ of all four languages -- one with the features of final consonants, the other with the features of initial consonants. (d) Cineradiography always shows two phases in the articulation of geminated /n/, /l/, /s/. The first phase is marked by consonant anticipation and weak tongue pressure, the second by vowel anticipation and increased tongue pressure. English emphasizes the first phase much more than the other languages. (e) Perceptual tests confirm that consonant duration is a major cue for the perception of gemination, and suggest that German, Spanish, and French ears require a longer hold than American ears in identifying a consonant as geminated, across word boundary. (f) Perceptual tests also indicate that the intensity of the arresting and releasing formant-transitions is a minor cue for the perception of gemination -- a cue liable to be overriden by the duration cue.

(B) Within word boundary, gemination is investigated in French contrasts such as Courait /kurɛ/ vs. Courrait /kurrɛ/, in Spanish contrasts such as Caro /karo/ vs. Carro /karro/, both compared with initial /r/ in Rama /rama/, [rrama], and in German contrasts such as Star /ʃtɑr/ vs. starr /ʃtarr/, by the same techniques as across word boundary.

Duration ratios between geminate /rr/ and single /r/ are found to be widest in Spanish (3.1 to 1, or 3.8 flaps to 1). In French and German they are comparable to those of /n/, /l/, /s/ (1.8 to 1). Signs of two phases, a final /r/ phase and an initial /r/ phase, in the amplitude displays and the motion picture x-rays, are clear in the French /rr/, not so clear in the Spanish /rr/, and not visible at all in the Spanish initial [rr] or the German final /rr/, which seem to behave like long /r/'s rather than like geminate /r/'s. In Spanish, therefore, /rr/ is essentially distinguished from /r/ by a much longer duration; in French by a longer duration, a two-phase articulation, and a narrower stricture; in German, by a longer duration and a more consonantal articulation (narrower stricture, stronger uvular trills), the final single /r/ being a particularly vocalic glide.

This study includes four tables of duration data comparing the geminate consonants with the single-final and single-initial consonants in each of the four languages. It also includes nine figures, one of the intensity variations of geminate and single consonants, one of the artificial-speech patterns used to test by ear the effect of varying the consonant duration, and seven of x-ray frames comparing articulatory sequences of single and geminate consonants.