https://doi.org/10.3176/lu.1998.3.16

MAIJA PIENIMÄKI (Turku)

FINNISH AND ENGLISH APPROXIMANTS: THE ACOUSTIC CONTINUUM

The approximant continuum [u-w-v-f-l-r] can most accurately be studied with controllable synthetic stimuli. Consequently, four experiments each containing six variants and two additional triplets were created. In experiments proper, only one parameter setting was varied at a time, which enables the isolation of the crucial aspects in the identification results. Three groups — two consisiting of Finns and one of native speakers of English — listed to the 30 synthesised utterances twice and made the forced choice between the alternatives *uii, wii, vii, fii, lii, rii* and the aim was to study the effect of the mother tongue system in the identification of various phones from the acoustic continuum. Despite the fact that the results still remain in the preliminary level, there seem to be some consistent differences between the answers given by the group consisting 13-year-old Finns and native speakers of English. However, the main objective of the present paper is to introduce the manners in which the synthesis was constructed.

1. Introduction

The terminology utilised to describe the group of phones consisting of [j], [w], [v], [l] and [r] varies from one linguist or phonetician to the next. This diversity may be caused, firstly, by the traditions in the study of individual languages — e.g. in connection with the Finnish phonology, the term *semi-vowel* is often used, whereas the French tradition appears to favour *semi-consonants* instead (Malmberg 1968 : 84). Secondly, the point of view adopted by the phonetician has an impact on the conventions used; for instance, P. Ladefoged (1975) justifies the choice of the label *approximant* with the articulative gestures involved. In addition, this terminology is also utilised by I. Morris-Wilson (1992 : 105—120), but a further division is made into lateral and central approximants. The name *glide* is used by several authors (viz. Lehiste, Peterson 1961; Chomsky, Halle 1968; Gimson 1962), who also mentions the name *frictionless continuant*. Altogether, there is a considerable amount of variation in the discussions on the group of sound in question. This does not, however, imply that the existence of this group would not be justifiable. Articulatorily, all these phones consist of a rapid movement of articulators, that is, the tongue or the lips. At the acoustic level this results in clear transitions of the formants, which act as the most distinctive perceptual cues for the recognition of these phones.

The objective of this study is to investigate the impact of the mother tongue in dividing the physical world of sounds into various categories. In the background there are several inter-related theories about the role of the first language. The tradition of contrastive analysis — outlined in R. Lado's famous work "Linguistics Across Cultures" (1957) — emphasised the p r e d i c t a b i l i t y of the errors in a situation of foreign language learning on the basis of the d i f f e r e n c e s between the category boundaries in the two systems. P. K. Kuhl (1993), on the other hand, concentrates on the role of the p r o t o t y p e s in the category centres; these prototypical variants — viz. Native Language Magnets — seem to pull the other members of the category towards the centre. The category prototypes, however, may differ in the first and the second language, and, consequently, learning difficulties occur. In the present study an attempt is made in order to reveal some of the linguistic factors, which are of importance in the identification of sounds from the acoustic continuum [u-w-v-f-l-r] in two languages, namely Finnish and English. In order to investigate this, controllable stimuli are required. For this reason, thirty synthetic syllables were created and played to three groups of informants, who were instructed to make the forced choice between the alternatives *uii, wii, vii, fii, lii* and *rii*. The first group consisted of 13-year-old native speakers of the Finnish Turku dialect[1]; the second listening test was done with first year students from the department of English at the University of Turku; and finally, some native speakers of English were asked to participate as well[2]. The results presented in this paper remain in the preliminary level due to the small amount of informants belonging to the third group. In this study, the emphasis is on the creation of the synthesis, not on the actual results. Consequently, only some preliminary results shall be given from two groups, namely Group 1 and Group 3.

2. The synthesis

The listening test consists of four basic series of six utterances and two additional triplets of syllables. The word initial context was chosen with a long high front vowel following immediately (the three lowest formants are at the frequencies of 250 Hz, 2400 Hz and 2900 Hz). The steady state vowel begins from the point of time 200 milliseconds (ms) and has the duration 290 ms. Consequently, the changes of interest take place during the first 200 ms. The approximants preceding the vowel all have the fixed value of 250 Hz for the first formant. Also, for the sake of simplicity, the second and the third formant always move together; these resonances are either set on the l o w alternative (F2: 600 Hz, F3: 1600 Hz) or the h i g h level (F2: 1600 Hz, F3: 2400 Hz). As a result, there remains four parameters to be studied: the duration of voicing, the duration of the transitions, the frequency of the loci for F2 and F3 and the effect of friction. Due to the fact that the transitions are of central importance in this investigation, there was no sense in including the approximant [j] in the synthesis; the sound in question hardly contains any glidings of the formants in the chosen phonetic environment. All the syllables were created using the Hcc Klsyn88 cascade/parallel synthesiser in the Center of Cognitive Neuroscience at the University of Turku.

[1] This dialect was chosen due to the existence of the voiceless fricative, which may be missing altogether from some local variants of the Finnish language.

[2] The native speakers of English represent several variants of the language, since recruiting only speakers of one particular dialect proved to be a nearly impossible task.

2.1. Experiment 1: The duration of voicing

In the first experiment three points of time are chosen for the initiation of voicing: the longest duration beginning at 0 ms with 40 dB and rising to 60 dB by 50 ms, the intermediate alternative starts with 40 dB at 50 ms and reaches 60 dB by 100 ms and, finally, the shortest duration has the value of 40 dB at 100 ms and 60 dB by 150 ms. The first three utterances are synthesised by using the l o w settings for the second and the third formant, whereas the syllables 4—6 have the h i g h values. Consequently, six different utterances result. The most crucial point is that the differences between the syllables are minimal in the sense that only one parameter value is varied at a time; therefore, also the formant transitions are kept constant with the steady state values continuing until 100 ms after which a steady glide towards the resonance frequencies for the adjacent vowel begin. It should, however, be noted that — due to the built-in characteristics of the synthesis programme — the transitions are audible only when the voicing amplitude is also present.

The role of the duration of voicing is expected to be of significance in the identification of sounds as speakers of two native languages listen to utterances. When the l o w values are chosen for F2 and F3 the resulting phone should be close to the Finnish [w] — already G. Fant showed that the protrusion and rounding of lips lowers the formants (1960 : 75). In addition, the increase in the duration of the first segment is expected to change the perceptual identification from [w] to [u]. Furthermore, the Finnish phonology lacks the phoneme /w/, which is only an allophone of the unit /u/. For speakers of English, however, the alternative *wii* ought to prove to be more attractive. The case when the h i g h setting for F2 and F3 is chosen, the resulting sound is close to the Finnish frictionless [v] approximant. In English this phone is clearly a fricative, which should be reflected in the answers. As a result, a group of Finns is expected to give more answers of the type *vii* in comparison to the English speaking group.

These tendencies appear to be present in the preliminary results of the experiment. When the l o w setting prevails and the point, where the voicing amplitude begins is at 0 ms Group 1 gives 41/50%[3] of the aswers as *uii*, but as the shortest voicing is chosen this alternative gets only 12/15% of the answers and the alternative *wii* is favoured. The English speaking group exhibits the same tendency, since all the subjects give the answer *wii*, when the minimal duration of voicing prevails, but the alternative *uii* is favoured by 43% of the informants as the longest duration of voicing is chosen (the alternative *wii* is still the most attractive one). With the h i g h F2-F3 settings a rather surprising result is found, since also the native speakers of English seem to favour the alternative *vii* both with the long and the short duration of voicing.

2.2. Experiment 2: The starting point of the transition

In this experiment the aim is study the effect that the starting point of the F2-F3 transitions has on the perceptual level. To begin with, the voicing amplitude is set to be similar to the intermediate position in the first experiment, in other words, the voicing amplitude starts with 40 dB at 50 ms and rises to 60 dB by 100 ms. Three points of time are chosen, namely 100 ms, 140 ms and 180 ms, and each of these is experimented both with the l o w and the h i g h F2-F3 loci values. As in the pre-

[3] This marking shows both the percentage figures as the utterance is heard for the first and the second time.

vious case in Experiment 1, only one parameter value is experimented with at a time and six utterances are created. An interesting aspect is connected with the fact that in this type of synthesis, the durations of the steady state and the transitional phase are obviously inter-related: as the transition is set to begin later, the duration of the preceding steady approrimant increases. This, however, is not the case in natural speech, where both of these parameters can be varied independent of each other. On the basis of the example words uttered by both a Finnish and an English informant it seems that, altogether, the English approximants may be of shorter duration. In the case of the Finnish approximants, it is possible to measure at least a brief period of a steady state phase; in contrast, the English variants appear to consist only of a transitional period.

In the preliminary results the most attractive answer alternatives remain the same in both the Finnish and the English speaking groups. However, what is interesting is the fact that answers of the type *rii* are also present for the group of native speakers of English, when the l o w setting prevails and the transition begin at the intermediate or late point of time. This alternative is not present in the answers of the Finnish group. Finns, for their part, give more *wii* answers at the expense of *vii* as the duration of the steady state period decreases and the h i g h setting is chosen. This is not the case with the English informants, who favour the alternative *vii*.

2.3. Experiment 3: The period of voiced friction

This experiment is devoted to the investigation of the impact of voiced friction. Again, both the l o w and the h i g h values for F2 and F3 are included. To begin with, three moments are chosen as the starting point for voicing: firstly, the duration of voicing is long and begins at 0 ms with 40 dB and rises to 60 dB by 50 ms, secondly, the intermediate alternative begins with 40 dB at 50 ms and reaches 60 dB by 100 ms, and thirdly, the shortest duration of voicing is chosen to begin at 100 ms with 40 dB and 60 dB by 150 ms. In this manner it is possible to see whether native speakers of Finnish and English react differently to the amount of friction. In Finnish, the sound [f] is the only phoneme, which could be considered as being a natural answer, since the system lacks the voiced fricative altogether. The sound [v] can sometimes be heard as an allophone of [f] in an inter-vocalic context. English, however, has both the voiced and the voiceless variants as phonological units.

This experiment gives clear results that indicate importance of the effect of voiced friction. When the l o w F2-F3 settings prevail, Finns favour the answer *fii* when voicing is either intermediate or short in duration, but the alternatives *wii* and *vii* become more dominant as the voicing is initiated at the earliest possible time. With the h i g h setting, Finns give various responses. For native speakers of English — with the l o w values —, the alternative *fii* is chosen in 50% of the cases when the voiceless period is maximally long, but when the voice is set on at the intermediate or the earliest possible point, the informants clearly favour the type *vii*. The h i g h values for F2 and F3 result in the alternative *vii* in almost 100% of the cases with each of the three durations of voicing.

2.4. Experiment 4: Voiceless friction

Experiment 4 studies the effect that the voiceless friction has on the identification. This is, once again, done both with the l o w and the h i g h values for the second and the third formant. The full amplitude of voicing is set to begin at the latest point of time in Experiment 3, viz. the pulse starts at 150 ms. As in the previous exper-

iment, three points of time were chosen for the commencing of the friction noise: in the first place, the amplitude has the value 40 dB at 0 ms and it reaches 60 dB by 50 ms, secondly, the amplitude begins at 50 ms with 40 dB and rises to 60 dB by 100 ms, and finally, 40 dB commences at 100 ms and the full value is present at 150 ms. Here, the role of the loci for the formants 2 and 3 may be of minor importance due to the fact that the transitions are audible only when voicing begins, namely at 150 ms. The setting where the fiction is of minimal duration may be problematic for both groups, since the fricative phone [f] ought to contain clear friction.

For the group of Finns, the most attractive alternative is the answer *fii*, but the percentage descends from 94/88% through 91/79% to 74/76 % as the duration of friction with the l o w F2-F3 values becomes shorter. This consistency is not repeated with the h i g h setting, in which case the alternative *fii* is considered to be the best alternative in each utterance, the percentage with the intermediate duration of friction exhibiting the clearest favouring. Also for Group 3, the answer *fii* is chosen most often with both of the F2-F3 values. However, the alternative *wii* is present when the F2 and F3 have the l o w values, but the answer type *vii* is important as the h i g h setting prevails. In the case, where the friction has the shortest duration, this alternative is even more dominant (with 50/83 %) than the answer *fii*.

2.5. Experiments L and R

These two triplets form a less systematic series in comparison with the Experiments 1—4. The unfortunate fact is that in these syllables more than one parameter is varied at a time. However, the aim was to synthesise variants, which could be identified differently by native speakers of two languages. In the case of the approximant [l], the synthesis is based on K. Wiik's (1966) analysis of the Finnish and English laterals. In each of the variant, the voicing amplitude is set to begin at 100 ms with 45 dB, rising to 50 dB by 160 ms and reaching 60 dB at 200 ms. The loci for the first three formants are varied: firstly, F1 is at 420 Hz, F2 is at 900 Hz and F3 has the frequency 2500 Hz; in the second place, F1 is at 400 Hz, F2 at 1000 Hz and F3 is at the level of 2100 Hz; and thirdly, F1 starts from 380 Hz, F2 from 1200 Hz and F3 has the value 2500 Hz. These frequencies are steady until the first 180 ms after which the formants glide to the appropriate frequencies for the following vowel. The synthesis of the approximant [r] — which is not a tremulant Finnish type of a phone — is even more unsystemantic and experimental than Experiment L. In all of the variants, F1 is fixed at the level 270 Hz until the moment 160 ms, when it descends to 250 Hz. The settings for the voicing amplitude are similar to those explained in connection with Experiment L and, consequently, the syllables actually start only from 100 ms. In the first utterance, F2 is at 1000 Hz at 100 ms and rises to 2000 Hz by 160 ms after which it continues to ascend to the frequency of the adjacent vowel, F3 changes at the same points of time from 1200 Hz to 2500 at 160 ms. The second utterance differs from the first one the with the initial loci for F2 and F3, which have the values 1100 Hz and 1500 Hz, respectively. In the third syllable, F2 starts from 1200 ms and rises to 2000 Hz by 160 ms; F3, on the other hand, begins from 2000 Hz and reaches 2500 Hz by 160 ms.

The preliminary results reveal that these triplets are not identified as consistently as the Experiments 1—4. However, what is interesting is that the darkest variant of the sound [l] is identified as *uii* by 41/47% of the Finnish informants, which follows the ideas presented by K. Wiik (1966 : 17). The clearer the [l] phone is, the less Finns favour the alternative *uii*. In the case of the approximant [r] native

speakers of English give more answers of the type *rii*, whereas Finns have inconsistent results with *wii* and *vii* as the most attractive alternatives.

3. Final remarks

Altogether, so far the most rewarding aspect of this study is connected with the creation of the synthetic utterances due to the fact that the results from the group of native speakers of English are still inconclusively. This contrastive point of view shall, however, be emphasised in the later stages of the investigation, since these result may in some degree shed light to the continuing problem of the possible differences between the identification of the physical world of sounds by native speakers of various languages. The fact that the synthesis of the first four experiments is conducted in a systematic manner enables one to concentrate on the actual individual parameters that may have the crucial impact on the classificational differences between the three groups of informants.

REFERENCES

Chomsky, N., Halle, M. 1968, The Sound Pattern of English, New York.

Fant, G. 1960, Acoustic Theory of Speech Production, The Hague.

Gimson, A. C. 1962, Introduction to the Pronunciation of English, Bristol.

Kuhl, P. K. 1993, Developmental Speech Perception. — Implications for Models of Language Impairment, 248—263.

Ladefoged, P. 1975, A Course in Phonetics, New York.

Lado, R. 1957, Linguistics Across Cultures, Michigan.

Lehiste, I., Peterson, G. E. 1961, Transitions, Glides and Diphthongs. — Journal of the Acoustic Society of America, 268—277.

Malmberg, B. 1968, Phonétique francaise, Lund.

Morris-Wilson, I. 1992, English Segmental Phonetics for Finns, Loimaa.

Wiik, K. 1966, Finnish and English Laterals, Turku.