Studies of the voice source

Mártony, J.

journal: STL-QPSR
volume: 6
number: 1
year: 1965
pages: 004-009

http://www.speech.kth.se/qpsr
Further voice source studies (1)(2) involving a larger speech material has been undertaken. The purpose was to obtain data on several different talkers and to study how the source spectrum of a vowel may depend on articulatory factors.

Speech material

The speech material consisted of the eleven long Swedish vowels /i:, e:, æ:, y:, ø:, ø:, u:, o:/ spoken in the contexts h-d or h-rd. Only /æ/ and /ø/ were spoken in the h-rd context. These vowels are positional variants of /e:/ and /ø:/, respectively, and occur exclusively before r. The material was read by twelve male talkers whose ages were between 19 and 45 years. They maintained normal subjective speaking level and articulation rate. There was a pause of about 2 sec between the utterances.

Method

The recorded material has been analyzed on the 51-channel spectrum analyzer set to provide harmonic analysis. The time point of a spectrum sample was chosen in a stationary part of the vowel. Amplitudes of each harmonic partial were measured from these data and compared with those synthesized according to our standard model. The difference values numerically calculated by means of a computer constitute a normalized voice source spectrum matching method formant bandwidths of the model are made a fixed function of the formant frequency and the higher pole correction is a function of $F_4$.

Remarks on errors in the measurements

In general the voice source (v.s.) spectra obtained from the front vowels give more even results. Those derived from back vowels show certain irregularities and deviations. These deviations can have several reasons. The back vowels /u/ and /o/ have very low levels in the medium frequency region and errors...
Fig. I-B-1. Mean voice source spectrum for speaker A derived from the vowels /i:, e:, e:, y:, æ:, u:, o:/ and normalized with respect to the -12 dB/oct voice source.
Fig. 1(b)-2. Mean voice source spectrum for speaker B derived from the vowels /i, ɪ, y, å, æ, u/ and normalized with respect to the -12 dB/oct voice source.
because of noise or distortion can easily arise. The vowel /u:/ is often nasalized and since the method has been developed for a non-nasalized model the deviation in the filter function shows up as abnormalities in the V.S. spectra. At the sound recordings the microphone was placed about 50 cm from the speaker's mouth. Direct sound radiation from the larynx or pharynx through the soft tissues of the neck are probably different for front and back vowels and could possibly influence the measured V.S. spectra.

Measurements made by Fujimura and Lindqvist (3) show that the first formant bandwidth for extreme low formant frequencies is higher than indicated by the experimental data of Fant (4) which were used for the present bandwidth approximations. There should also be taken into account the specific deviations of formant bandwidths of a vowel from the statistical average. Any deviation from a perfect constant phonation and articulation of the subject at the interval of sampling may produce frequency modulation effects causing an apparent broadening of formant peaks. In some of the V.S. spectra discussed below there are irregularities which may depend on errors in the formant bandwidths assumed for the inverse filtering.

Results for individual talkers

Figs. I-B-1 through I-B-12 show mean values of V.S. spectra derived from different vowels for 12 speakers. In these figures the V.S. spectra are normalized with respect to the -12 dB/oct voice source, i.e. showing amplitude versus frequency in excess of the -12 dB/oct standard source.

Speaker A

Fig. I-B-1 shows the mean V.S. spectrum derived from the vowels /i:, e:, e:, y:, \, u:, o:/\/. This spectrum has one minimum at 0.8 kc/s and a second one at about 2.6 kc/s. The range of the deviation from the mean curve is about \( \pm 7.5 \) dB.

Speaker B

Fig. I-B-2 shows the speaker's mean V.S. spectrum of the vowels /i:, e:, y:, \, o:, u:/\/. The V.S. spectra of the
remaining vowels display some irregularities. The spectrum of Fig. I-B-2 has a minimum at about 0.9 kc/s, which is somewhat smoothed out because the minima of the individual v.s. spectra are varying in frequency between 0.6 and 0.9 kc/s. The spectrum falls faster than -12 dB/oct above 1 kc/s. The range of the deviation from the mean spectrum is about ±5 dB except in the region of the minimum where it is +5 to -10 dB.

Speaker C

Fig. I-B-3 shows the mean v.s. spectrum derived from the front vowels /i:, e:, ë:, æ:, y:, ŋ:, ɐ:, a:/ The back vowels on the other hand had apparent irregularities in the middle part of the spectrum. The spectrum of this speaker is low-frequency emphasized and slopes off at a rate faster than -12 dB/oct at higher frequencies. The maximum deviation from the mean is ±5 dB except in the region 0.6 to 1.2 kc/s where it is +5 to -12 dB because of the voice source zero.

Speaker D

The v.s. spectra of this speaker show consistent differences with respect to the close-open vowel category (see Fig. I-B-4). Thus the v.s. spectra of the vowels /ɛ:, ə:, ø:, æ:, u:, œ:, ə:/ are characterized by an over-all slope of -12 dB/oct, a minimum between 0.5 and 1.5 kc/s, and another minimum at 2.4 kc/s. The mean of the class of close vowels /i:, e:, y:, ɐ:/ has a more steeply falling v.s. slope and a more definite minimum at 1 kc/s. The maximum deviation from the mean v.s. spectrum of both classes of vowels is ±5 dB.

Speaker E

The mean v.s. spectrum derived from the front vowels /i:, e:, ë:, æ:, y:, ŋ:, ɐ:/ is shown in Fig. I-B-5. This spectrum does not have any clear minimum. Its slope is greater than -12 dB/oct. The v.s. spectra of the vowels /u:/ and / œ:/ show a quite different character. They have a less dominating low frequency region and a very definite minimum at 0.8 kc/s. The range of the v.s. deviation from the mean of this speaker is ±10 dB, a value which is large compared with those obtained from the other speakers.
Fig. I-B-3. Mean voice source spectrum for speaker C derived from the vowels /i:, e:, ə:, y:, ʊ:, ø:/ and normalized with respect to the -12 dB/oct voice source.
Fig. I-B-4. Mean voice source spectra for speaker D. Solid line: voice source spectrum derived from the vowels /ɛ:, e:, ɨ:, ʊ:, ʌ:, a:/. Dotted line: voice source spectrum derived from the vowels /iː, eː, ɨː, ʊː, ʌː/. The spectra are normalized with respect to the -12 dB/oct voice source spectrum.
Fig. I-B-5. Mean voice source spectrum for speaker E derived from the vowels /i:, e:, e:, a:, y:, ə:, θ:/ and normalized with respect to the -12 dB/oct voice source.
Speaker F

The v.s. spectra of this speaker are of two different types (see Fig. I-B-6). The mean v.s. spectrum derived from the front vowels /i:, e:, ε:, ϕ:, æ:/ falls faster than -12 dB/oct. The mean v.s. spectrum of the back vowels /u:, o:, a:/, on the other hand, has a relatively low level at frequencies below 1 kc/s. This effect may be partly dependent on the circumstances discussed in the introduction. The range of the deviation from the mean is ±8 dB in both types of spectra.

Speaker G

Fig. I-B-7 shows the mean v.s. spectrum of the front vowels /i:, e:, ε:, æ:, y:, ϕ:, æ:, ω:/, A minimum which is not too prominent is located at 0.9 kc/s. The over-all slope of the spectrum is somewhat greater than -12 dB/oct. The deviation from the mean value is not more than ±5 dB.

Speaker H

Fig. I-B-8 shows the mean v.s. spectrum of the vowels /i:, e:, ε:, y:, ϕ:, æ:, ω:/, There is a dominant minimum at 0.8 kc/s. The deviation from the mean values is quite small, about ±3 dB, except in the region of the second zero at 2.0 kc/s, where it is +5 to -7.5 dB.

Speaker I

Fig. I-B-9 shows the mean v.s. spectrum derived from the front vowels /i:, e:, æ:, y:, ϕ:, æ:, ω:/, There is a dominant zero at 0.8 kc/s and a second less prominent zero at 2.1 kc/s. The maximum deviation from the mean values is about ±5 dB. The mean v.s. spectrum derived from the back vowels is about 2-3 dB less in the low frequency region (below 1 kc/s), and the range of the deviation from the mean values is less than ±2.5 dB in this region.

Speaker K

The v.s. spectra of this speaker show several different characteristics (see Fig. I-B-10). The mean v.s. spectrum
Fig. I-B-6. Mean voice source spectra for speaker F. Solid line: voice source spectrum derived from the vowels /i:, e:, e:, o:, a:/.
Dotted line: voice source spectrum derived from the vowels /u:, o:, a:/.
The spectra are normalized with respect to the -12 dB/oct voice source.
Fig. I-B-7. Mean voice source spectrum for speaker G derived from the vowels /i:, e:, æ:, y:, ø:, ø:/ and normalized with respect to the -12 dB/oct voice source.
Fig. 1-8-8. Mean voice source spectrum for speaker II derived from the vowels /i:, e:, ɛ:, ʊ, ɔ, ʌ, w, ɔɪ/ and normalized with respect to the -12 dB/oct voice source.
Fig. I-B-9. Mean voice source spectrum for speaker I derived from the vowels /i:, e:, z:, y:, ø:, æ:, ø:/ and normalized with respect to the -12 dB/oct voice source.
Fig. I-B-10. Mean voice source spectra for speaker K.
Solid line: voice source spectrum derived from the vowels /i:, y:, ɪ:/.
Dotted line: voice source spectrum derived from the vowels /ʊ:, u:, ə:/.
Dashed and dotted line: voice source spectrum derived from the vowels /e:, ɛ:, æ:/.
The spectra are normalized with respect to the -12 dB/oct voice source spectrum.
derived from the vowels /i:, y:, ø:/ and the mean v.s. spectrum derived from the vowels /e:, æ, æ:/ both fall at a rate greater than -12 dB/oct in the higher frequency region. In the former mean spectrum the zero at 0.6 kc/s is more definite. The mean v.s. spectrum derived from the vowels /æ:, u:, ø:/ has a definite zero too, but its approximation to the -12 dB/oct standard is not equally good. The speaker has a certain instability of phonation. The maximum deviation from the mean values in all the three cases is ±3 dB except at about 2.5 kc/s where it is ±5 dB.

Speaker L

Fig. I-B-11 shows the mean v.s. spectrum derived from all the vowels, i.e. /i:, œ:, æ, y:, æ, ø:, æ, u:, o, ø/.

In the v.s. spectra of some vowels there is a definite minimum at 1 kc/s and for the others there is no minimum. The maximal deviation from the mean v.s. spectrum is ±7.5 dB.

Speaker M

Fig. I-B-12 shows the mean v.s. spectrum derived from the vowels /i:, œ:, æ, æ, y:, æ, ø:/.

There is no definite minimum in the spectrum except at 2.5 kc/s. The maximal deviation from the mean v.s. spectrum is ±5 dB, in some parts only ±2.5 dB.

Comparison of the v.s. spectra for different speakers

The numerical spectrum matching was more successful with front than with back vowels. The analysis of back vowels is difficult owing to

1) possible nasalization (open vowels),
2) possible errors in formant bandwidths,
3) low signal to noise ratio (at medium frequencies),
4) radiation through the neck.

Fig. I-B-13 shows mean v.s. spectra derived from different vowels. All the v.s. spectrum curves show a more quickly falling spectrum than the -12 dB/oct normal source and there appear to be two spectral minima at about 0.7-0.9 kc/s and 2.5 kc/s.

The mean v.s. spectra of the speakers can be classified into three groups.
Fig. I-B-11. Mean voice source spectrum for speaker L derived from the vowels /i:, e:, æ:, u:, ə:, a:, o:, ɔ:/ and normalized with respect to the -12 dB/oct voice source.
Fig. I-B-12. Mean voice source spectrum for speaker M derived from the vowels /i:, e:, e:, e:, y:, st, m:/ and normalized with respect to the -12 dB/oct voice source.
Fig. I-B-13. Mean voice source spectra for different speakers. Dotted line: group 1) slight fall. Solid line: group 2) normal fall. Dashed and dotted line: group 3) steep fall. The spectra are normalized with respect to the -12 dB/oct voice source.
Fig. I-B-14. Mean voice source spectrum derived from all the vowels of all the speakers. The upper and lower lines show the deviation from the mean.
1) slight fall (-12 to -13 dB/oct),
2) normal fall (about -15 dB/oct),
3) steep fall (about -18 dB/oct),

see Fig. I-B-13. The maximum deviation around the mean values is ±8 dB, in some local regions ±12 dB. The differences between groups 1, 2, and 3 are significant.

Comparison of v.s. spectra for different vowels

Fig. I-B-14 shows the mean v.s. spectrum for all the vowels and all the speakers. The spectrum is falling by about -15 dB/oct and contains one minimum at about 0.7 kc/s and another one at 2.4 kc/s. The maximum deviations from the mean are shown too. These deviations are in some cases as large as 15-20 dB. The mean v.s. spectra derived from all the speakers for

1) unrounded front vowels,
2) rounded front vowels, and
3) back vowels,
do not show significant differences from each other. The v.s. spectrum for the back vowels deviates a little from the others probably for the reasons discussed earlier.

Summary

The present investigation shows certain variations of the v.s. spectrum with respect to speaker. A systematic variation of the v.s. spectrum with the articulation could not be shown except for individual speakers. These results are in agreement with our earlier investigations (1)(2) as well as with a recent investigation of Carr and Trill (5), which was concerned with English vowels spoken by American subjects.

References: