UNIVERSITY OF CAMBRIDGE

CNBH, Physiology Department, Cambridge University
www.mrc-cbu.cam.ac.uk/cnbh

CNBH

# Size discrimination in CV and VC English syllables. P63


David. T. Ives and Roy. D. Patterson

tim.ives@mrc-cbu.cam.ac.uk roy.patterson@mrc-cbu.cam.ac.uk


## 1. Introduction

Size information is contained in sound. In humans, this information is conveyed by the vocal tract.

A wide range of different vocal tract lengths exist in the population but humans are able to distinguish the source (speaker) from the content (speech).

This suggests that the size and shape information about the vocal tract can be segregated [Irino and Patterson (2002)].

Experiments by Smith and Patterson (2004) used vowel sounds to determine the range of speaker sizes over which the identification of vowel type and discrimination of speaker size was possible.

We extend the size discrimination experiment using a larger database of speech sounds and show that performance increases compared with vowel sounds.

We also look at size discrimination performance for different speech sounds and show there is a small improvement for consonant-vowel speech as opposed to vowel-consonant speech.

## 2. Experimental procedure

The experiment is a 2AFC paradigm design in which the subjects are required to discriminate the scale between sequences of utterances.

Stimuli are speech syllables resynthesised by STRAIGHT [Kawahara (1997)], also P-centre corrected.

STRAIGHT is a high-quality vocoder it separates the VTL and GPR information in speech sounds and resynthesises the same utterance over a range of VTL and GPR values.

Speech is taken from a database created at the CNBH. The database contains 180 unique syllables and additional versions resynthesised at many different GPRs and VTLs.

Discrimination performance is measured at five points in the GPR-SER space (see Fig. 1), where SER is α 1/VTL.

At each point a discrimination is made between the test size and three larger and three smaller speakers. Level and pitch cues are roved.

Measurements are made for 6 groups: CV-sonorants, CV-stops, CV-fricatives VC-sonorants, VC-stops and VC-fricatives.

Six subjects, 40 repetitions for each point on a psychometric curve.

| | Sonorants | | | | | | Stops | | | | | | Fricatives | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CV's | ma | na | la | ra | wa | ya | ba | da | ga | pa | ta | ka | sa | fa | va | za | xa | ha |
| | me | ne | le | re | we | ye | be | de | ge | pe | te | ke | se | fe | ve | ze | xe | he |
| | mi | ni | li | ri | wi | yi | bi | di | gi | pi | ti | ki | si | fi | vi | zi | xi | hi |
| | mo | no | lo | ro | wo | yo | bo | do | go | po | to | ko | so | fo | vo | zo | xo | ho |
| | mu | nu | lu | ru | wu | yu | bu | du | gu | pu | tu | ku | su | fu | vu | zu | xu | hu |
| VC's | am | an | al | ar | aw | ay | ab | ad | ag | ap | at | ak | as | af | av | az | ax | ah |
| | em | en | el | er | ew | ey | eb | ed | eg | ep | et | ek | es | ef | ev | ez | ex | eh |
| | im | in | il | ir | iw | iy | ib | id | ig | ip | it | ik | is | if | iv | iz | ix | ih |
| | om | on | ol | or | ow | oy | ob | od | og | op | ot | ok | os | of | ov | oz | ox | oh |
| | um | un | ul | ur | uw | uy | ub | ud | ug | up | ut | uk | us | uf | uv | uz | ux | uh |

Table 1. Stimuli set showing categories of CVs, VCs and sonorant, stops and fricatives.

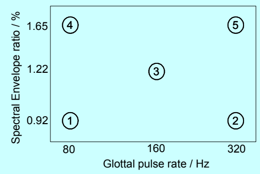


Figure 1. Measurement points on GPR-SER space.

## 3. Results

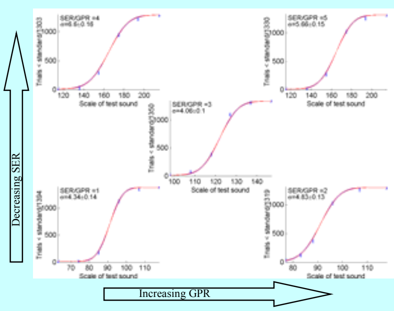


Figure 2. Psychometric functions for the five SER-GPR conditions, summed over all subjects and all syllable groups.

| Stimuli Group | SER/GPR Condition 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| All | 4.3±0.1 | 4.8±0.1 | 4.0±0.1 | 6.6±0.1 | 5.6±0.1 |
| CV | 4.0±0.1 | 4.9±0.1 | 3.9±0.1 | 5.9±0.2 | 5.3±0.2 |
| VC | 4.4±0.2 | 4.3±0.1 | 4.4±0.1 | 7.6±0.2 | 6.2±0.2 |
| Sonorant | 4.0±0.2 | 4.1±0.1 | 4.0±0.1 | 5.8±0.2 | 5.0±0.2 |
| Stop | 4.5±0.2 | 6.2±0.2 | 4.2±0.1 | 8.3±0.3 | 6.5±0.2 |
| Fricative | 4.2±0.2 | 4.3±0.2 | 3.7±0.1 | 5.7±0.2 | 5.3±0.2 |
| *Smith (2004)* | *10.5* | *17.2* | *6.6* | *10.6* | *9.3* |

Table 2. JND values for all syllable groups averaged over all subjects.

## 4. Analysis

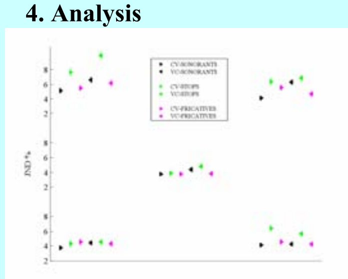

Figure 3. JND values for SER-GPR condition as a function speech group. SER-GPR groups arranged as in Figure 1.

SER-GPR=1,3: JNDs stable at 4% for all speech groups.

SER-GPR=2: JND increases slightly but more for stops.

SER-GPR=4,5: JNDs increase to 7% and 6%, for stops increase is more.

All speech groups except for stops and to a lesser degree VC's the JND values are similar within an SER-GPR condition.

**Compared to vowels:**

SER-GPR=3,4,5: 30% decrease in JND.

SER-GPR=1: 60% decrease in JND.

SER-GPR=2: 70% decrease in JND.

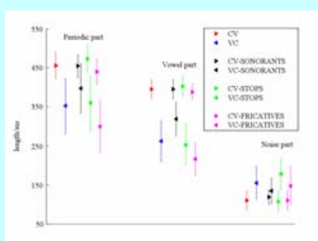


Figure 4. Duration of "periodic", "vowel" and "noise" components for each speech group (error bars show ±1 standard deviation).

Categorise syllable content into three components:

*(i) periodic component (including any vowel information),*

*(ii) vowel component,*

*(iii) noise component.*

CV's have longer periodic components than VC's and significantly longer vowel components.

Fricatives have the least periodic and vowel content followed by stops and then sonorants, which have the greatest.

Only a small variation in the duration of the noise part.

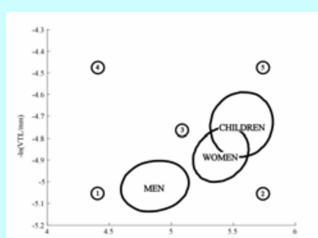


Figure 5. Vocal tract length and glottal pulse rate distributions of men, women and children [replotted from Peterson and Barney (1952)]. Numbered circles show the SER-GPR conditions used in the experiment.

The enclose 90% of the population.

VTLs and GPRs that we would encounter in everyday life.

Numbered circles show the SER-GPR conditions in the experiment.

Size discrimination possible outside the normal ranges of size.

Supports the argument of Smith et al. (2004): the auditory system includes a scale normalisation process.

## 5. Conclusions

It is possible to discriminate size using speech.

Experiments have shown JND values for size discrimination of between 4% and 6%.

Discrimination is good and consistent outside normal range of speech, which supports the argument for a scale transform in auditory system.

Improvement in JND values over vowel of between 30% and 70%.

Small performance increase for CV over VC, may be due to increased periodic content.

## References


*Irino and Patterson (2002). Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-Mellin transform. Speech Communication.*

*Kawahara, H. (1997). Speech representation and transformation using adaptive interpolation of weighted spectrum: vocoder revisited. Proceedings IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP '97) 2, 1303-1306.*

*Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H., and Irino, T. (2004). "The processing and perception of size information in speech sounds," (review process, submitted JASA).*

*Peterson, G.E., Barney, H.L. (1952) Control methods used in the study of vowels. JASA 24(2): 175-184.*



**Acknowledgements**

Research supported by the U.K. Medical Research Council (G9901257).