Irfan Umer Munir - Mohammad Bin Shahbaz

Categorical Perception

Abstract: *This paper introduces the phenomenon of categorical perception and establishes it by performing an experiment on the [s-ʃ] continuum in which the bandwidth of the fricative noise is varied gradually from a narrow band of [s] to a wide band of [ʃ]. The two fricatives are distinctively identified without any transitional ambiguity even though the variation is gradual.*

Keywords*: speech perception, speech synthesis, stimuli, frication noise, categorical perception.*

1. INTRODUCTION

Speech perception is the phenomenon in which humans interpret a speech sound stream. A common theory is that humans employ special decoding processes to convert the patterns of the sound stream back into a string of phonemes that define the word of the message (Pickett, 1999). An interesting point is that not all the sound streams are decoded in the same way. The mechanism of interpreting the speech sounds is different from that of non-speech sounds and similarly the mechanism of perceiving vowels is different from that of the consonants.

Categorical perception means that a continuous gradual change in some variable along a continuum is perceived, not as gradual but as an instance of discrete categories (www.coglab.wadsworth.com).

People perceive most stimuli continuously. For example, when you look at a rainbow, you see a smooth transition from black to white, as seen in Figure 1.

Figure1 Smooth perception.

Stimuli are not usually perceived categorically. An example of categorical perception is that either pure black, or pure white can be seen (nothing in between), as seen in Figure 2.

Figure 2 Categorical perception

This phenomenon is called "categorical" perception because instead of getting a percept that is ambiguous, you get a percept that perfectly matches an ideal example of a particular category. One thing that people seem to perceive categorically is speech (*Advanced Topics: Categorical Perception*). What is interesting about this is that even when the physical stimuli change continuously (as in Figure 1), people perceive it categorically (as in Figure 2).

2. LITERATURE REVIEW

To establish the phenomenon of categorical perception, experiments have been performed in which a series of CVs were synthesized by varying a particular acoustic feature in equal steps along a continuum. The listeners did not perceive a continuous change in the sounds; rather they found perceptual boundaries of the consonants and the stimuli fell in one of the discrete categories. Categorical perception was noted for dimensions of phonetic contrast such as place of articulation, voicing and manner of articulation. The role of place of articulation is described in Experiment 1 and the role of voicing is described in Experiment 2.

2.1 Experiment 1

In one early study, a continuum of stop-vowel syllables was constructed where the syllables differed only in the direction and extent of the second formant transition. This experiment was conducted by Liberman, Harris, Hoffman and Griffith in 1957 (Pickett, 1999). The F2 transitions ranged over a [b]-[d]-[g] continuum. When presented with this series of synthetic speech patterns, most listeners reported hearing all the stimuli as clear cases of “bay”, “day” or “gay”. Although the change is gradual, this is not the way it was perceived. Subjects regularly perceived the different stimuli as being instances of either of the three syllable types, \`ba', \`da' or \`ga'.

2.2 Experiment 2

Another synthetic continuum of stimuli was developed at Haskins Labs in which the voiced onset time was systematically varied. The experiment was conducted by Ambramson and Lisker in 1967 (Pickett, 1999). Both /b/ and /p/ are stop consonants. To produce these, lips are closed then opened, some air is released, and the vocal chords begin vibrating. The difference between /ba/ and /pa/ is the time between the release of the air and the beginning of the vibration. This is referred to as voice onset time or VOT. For /b/, VOT is very short; voicing begins at almost the same time as the air is released. For /p/, the voicing is delayed. When people were asked to identify these stimuli, they generally had no difficulty. The first few were identified as /b/ and the second few were identified as /p/. What was most interesting was how the middle items were identified. Unlike most other stimuli, people did not report hearing something that was a bit like /b/ and a bit like /p/. Rather, they reported hearing either /b/ or /p/.

2.3 Results of Experiments

In the first experiment mentioned above the place of the stops was changed by varying their F2 onset frequency and as a result the three stops were uniquely identified as labial, alveolar and velar stops. In the second experiment mentioned above the voicing of the two stops was changed by varying the onset of the first frequency, and in this experiment also the two stops were identified uniquely even though the voicing was changed gradually.

The experiments performed on categorical perception consisted of two major tests, the identification test and the discrimination test. In the identification test the stimuli were randomized and the listeners were asked to label them. In the discrimination test the listeners were asked to tell whether the two stimuli produced the same sound or not.

By performing these tests it was concluded that perception of an acoustic feature that varies along a continuum of equal steps is discontinuous. The most noteworthy point is that this discontinuity corresponds to the boundaries between the phoneme categories. It was concluded that a listener could better discriminate between sounds that lie on the opposite sides of a phoneme boundary than between sounds that fall within the same phoneme category. This conclusion was made by Liberman, Harris, Hoffman and Griffith in 1957 (Pickett, 1999).

The results of the identification test revealed that labeling of stimuli showed abrupt, steep slopes between adjacent phonemic categories and no change within the same phonemic category. The results of the discrimination test revealed that discrimination of stimuli from within the same phonemic category tends to be poor, whereas discrimination between two stimuli from different phonemic categories is good.

2.4 Categorical perception in animals

Two very major questions rise regarding categorical perception. The first is that does the phenomenon also hold for animals. The answer is yes, it does. Experiments were performed on animals and it was observed that when a continuum of stimuli was delivered to the animals they responded in the same manner for a number of stimuli but showed abrupt response after a series of stimuli (Pickett, 1999). Thus their response was not gradual, but abrupt after every few stimuli.

2.5 Categorical perception in vowels

Another interesting question is whether categorical perception holds equally for vowels and non-speech sounds as well or it is true for consonants of speech sounds only. Experiments performed on the vowels revealed that the perception of the vowels is continuous. Fry, Abramson, Eimas, and Liberman, in 1962, (Pickett, 1999) investigated a continuum that ranged from /□/ to /e/ to /□/ and found that discrimination among items from within the same vowel category was quite good. Thus, rather than being categorical, the perception of vowel contrasts was continuous. The cause of this non-categorical behavior of the vowels was stated to be their difference in the condition of articulation. In the production of consonants there are discrete motions that have to attain certain targets whereas for the vowels the tongue position can assume a large number of different positions for the production of the same vowel. Thus it was concluded that the degree to which categorization of perception occurs along the acoustics dimensions depend on whether the sounds are produced categorically (consonants) or non-categorically (vowels). Perception seems to correspond to the nature of articulation rather than the continuum of the acoustic features (HLSYN).

2.6 Categorical perception in non-speech sounds

Non-speech sounds also do not support the phenomenon of categorical perception, i.e. if the acoustic feature causing the variation in the stimuli was listened to in isolation (with all the other acoustic features deleted) a gradual change among the stimuli was observed. Thus we can conclude that speech perception is special and the decoding process involved in interpreting speech sounds is quite unique.

3. METHODOLOGY

The experiment we performed to establish the phenomenon of the categorical perception, was on the [ʃ -s] continuum in which the bandwidth of the fricative noise was varied gradually from a narrow band of [s] to a wide band of [ʃ] and then the identification test was performed on the stimuli formed. To perform this experiment the sounds of [sɑ] and [ʃɑ] were recorded. Using the software, Winsnoori version 1.3, the formants (F1, F2, F3 and F4) of the two sound streams were determined using an interval of 0.01 second. Taking the formant values, the two sound streams were synthesized using the software HLSYN (High Level Synthesizer). Besides the formant values, other acoustic features were also adjusted to synthesize more natural sound streams. HLSYN has a total of thirteen parameters (HLSYN). The parameters that we were concerned with are as follows:

- F0 forms the fundamental frequency of vocal fold vibration.
- F1, F2, F3, F4 form the first four natural frequencies of the vocal tract.
- ‘ag’ controls the area of glottal opening.
- ‘ab’ controls the cross sectional area of constriction formed by the blade of the tongue (Rafique, 2002).

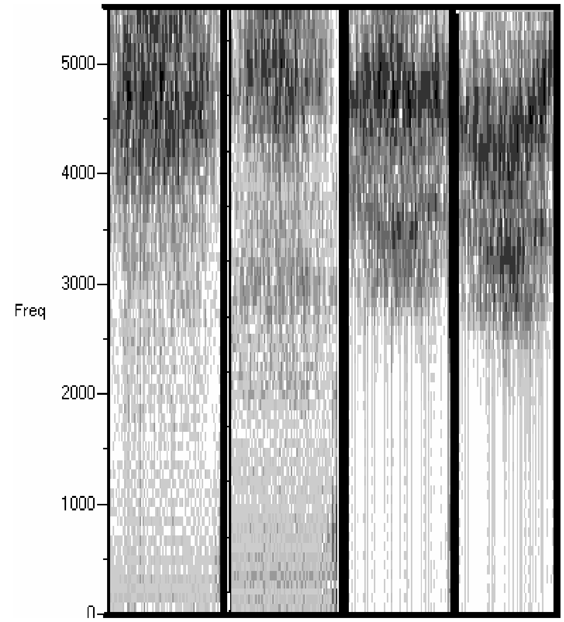
The values of F0 is almost same for all the consonants, therefore after studying different synthesized consonants, the same values were used in the synthesizing [sɑ] and [ʃɑ]. The values of 'ag' and 'ab' were also adjusted according to the instructions given in the guide of HLSyn for synthesizing voiceless fricatives (HLSYN). The default values of the rest of the parameters were taken as they did not impose any impact in the synthesis of [sɑ] and [ʃɑ]. In the experiment [sɑ] was synthesized having a noise level at 3800 Hz and [ʃɑ] was synthesized having a noise level at 1900 Hz.

Once the sound streams of [sɑ] and [ʃɑ] were synthesized, the next step was to synthesize the intermediate stimuli in which the bandwidth of the fricative noise was increased gradually. The increment in the voice was primarily controlled by two parameters: F2 and 'ab'. It was observed that F2 corresponds to the articulatory position and thus by gradually increasing it, the place of articulation was shifted from the alveolar fricative [sɑ] to the palatal fricative [ʃɑ]. The F2 for the first stimulus was 1600 Hz and that of the last stimulus that was 2600 Hz. The second parameter was 'ab' that controlled the tongue blade. In [sɑ] the front portion of the tongue was used whereas in [ʃɑ] the back portion of the tongue body was being used, thus the parameter 'ab' was gradually changed to reflect the intermediary tongue positions, as shown in Appendix B. The values of the other parameters were kept constant, as they did not play any significant role in increasing the noise level. Moreover the parameters of the vowel 'ɑ ' were kept the same for both the fricatives.

After synthesizing stimuli with a continuum of 10 steps, the stimuli were randomized and placed in sets; each set contained five random stimuli. Five such sets were made and ten listeners were asked to listen to all the sets of stimuli. The listeners had to listen to the stimuli and label each stimulus as either [sɑ], [ʃɑ] or as an ambiguous sound. The listeners that were chosen for this experiment belonged to the same age group (18-22 years) and all had normal hearing capacity. After all the observations had been obtained, a graph was plotted that depicted the obtained observations and showed whether the sound streams were heard categorically or continuously.

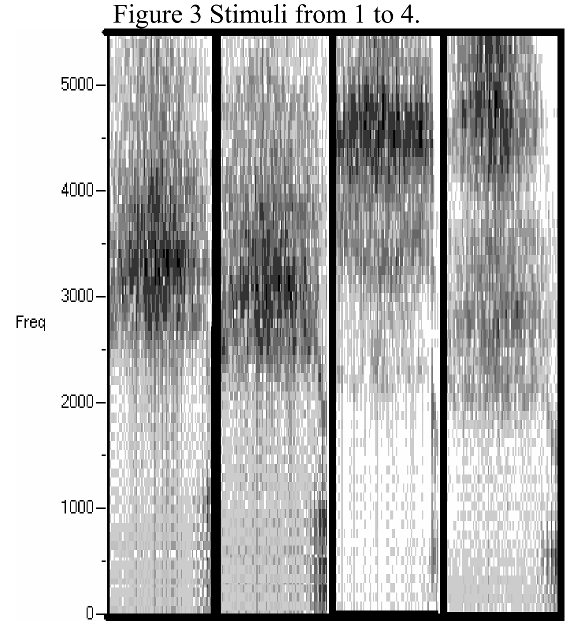
4. RESULTS

The results obtained from the listeners are shown in Appendix A. The results show that the first two stimuli are clearly heard as [sɑ] and the remaining 6 stimuli are heard as [ʃɑ]. Though some listeners found some difficulty in perceiving the intermediate stimuli but that was just because the stimuli were not very natural. However the results were substantial enough to support the existence of the phenomenon of categorical perception.



stimulus1 stimulus2 stimulus3 stimulus4

Figure 3 Stimuli from 1 to 4.



stimulus5 stimulus6 stimulus7 stimulus8

Figure 4 Stimuli from 5 to 8,

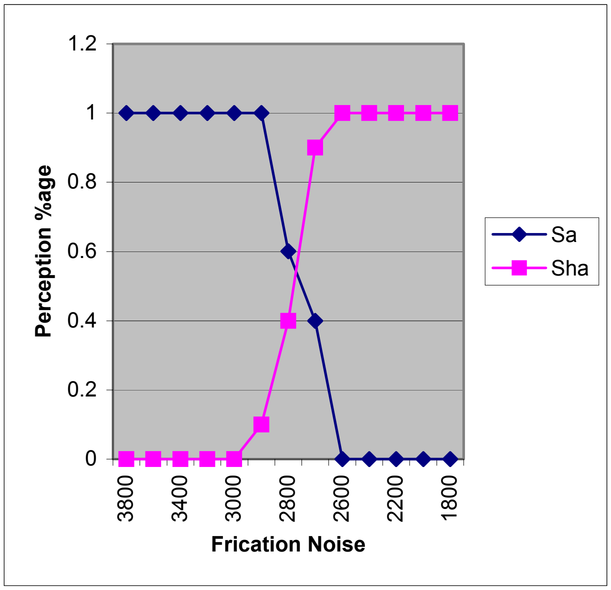


Figure 5 Perception of [s-ʃ] continuum.

The graph shown in Figure 5 forms a steep slope, showing an abrupt change at the phoneme boundary and no change within the same phoneme category.

5. DISCUSSION

There a number of very interesting observations that came forward while performing this experiment.

Firstly the bandwidth of the sound streams of [sɑ] and [ʃɑ] continuum was found to range from 3800 Hz to 1900 as is clear from Figure 3 and Figure 4. The bandwidth of the first stimulus is 3800 Hz and that of the eighth stimulus is 1900 Hz. However the documented range of the [s-ʃ] continuum is 3000 Hz to 2000 Hz.

Secondly, it was observed that the variation in F2 was proportional to the change in the bandwidth of the fricative noise. The variation in the 'ab' parameter that was controlling the position of the tongue did not bring a significant change in the noise level, i.e., a change in the 'ab' parameter was not apparent in the noise level of the stimuli; however it brought a considerable change in the perception of the sound.

It was found from the results obtained by the listeners that almost all of them categorized the first two stimuli as [sɑ] and the remaining six stimuli as [ʃɑ]. Thus the noise level that formed the phoneme boundary of the two sound streams was observed to be at 2750 Hz. The value of F2 at this noise level was found to be 2000 Hz. and the value of 'ab' was found to be initially at 30 and ending at 40 as shown in the parameter values of the stimuli in Appendix B. No one perceived the first two stimuli as [ʃɑ] and the last 5 stimuli as [sɑ]. Thus it was concluded that we could better discriminate between sounds that lie on the opposite sides of a phoneme boundary than between sounds that fall within the same phoneme category. Thus the experiment very clearly proved that the phenomenon of categorical perception for speech was true for the consonants.

6. SUMMARY

It is an intriguing concept that a continuous change in the acoustic features can be perceived as a discrete abrupt change. Our experiment has revealed results that conform to the phenomenon of categorical perception of speech sounds.

For future enhancements, this experiment can be extended to observe non-categorical perception of the non-speech sounds, i.e., if the values of F2 and 'ab', causing the variation in the stimuli, are listened to in isolation (with all the other acoustic features deleted) a gradual change among the stimuli should be observed.

REFERENCES

- Advanced Topics Categorical Perception (CP). University of South Ampton.
- Clark John and Yallop Collin. *An Introduction to Phonetics and Phonology.* Second edition 1996-99
HLSYN high-level speech synthesizer. User Interface Manual. Sensimetrics Corporation. Somerville Massachusetts.
- Pickett J.M. The Acoustics of Speech Communication. Published 1999.
- Rafique Mohammad Mustafa. Synthesis of /ikis/(21), /bias/(22). Akhbar-e-Urdu April-May 2002.
- www.coglab.wadsworth.com
- www.ling.umu.se/~rand/KatPer/index.eng.html

APPENDIX A

Listener 1

	[sɑ]	[ʃɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	5	0	
Stimulus 3	1	3	1
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 2

	[sɑ]	[ʃɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	4	1	
Stimulus 3	0	4	1
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 3

	[sɑ]	[ʃ	Ambiguous

		ɑ]	
Stimulus 1	5	0	
Stimulus 2	2	1	2
Stimulus 3	0	5	
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 4

	[sɑ]	[ʃ ɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	5	0	
Stimulus 3	0	5	
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 5

	[sɑ]	[ʃ ɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	3	0	2
Stimulus 3	0	4	1
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 6

	[sɑ]	[ʃ ɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	5	0	
Stimulus 3	1	3	1
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 7

	[sɑ]	[ʃ ɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	4	0	1
Stimulus 3	0	4	1
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

Listener 8

	[sɑ]	[ʃ ɑ]	Ambiguous
Stimulus 1	5	0	
Stimulus 2	3	1	1
Stimulus 3	0	3	2
Stimulus 4	0	5	
Stimulus 5	0	5	
Stimulus 6	0	5	
Stimulus 7	0	5	
Stimulus 8	0	5	

APPENDIX B

Parameter of stimulus 1

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	Ap
0.0	15.00	100.0	5.000	0.0	0.0	1500	350.0	1600	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	12.00	0.0	0.0	1700	409.0	1600	2750	3500	8.000	0.0	0.0
180.0	10.00	100.0	20.00	0.0	0.0	1000	468.4	1600	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	20.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 2

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	10.00	0.0	0.0	1500	350.0	1800	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	15.00	0.0	0.0	1700	409.2	1800	2750	3500	8.000	0.0	0.0
180.0	10.00	100.0	25.00	0.0	0.0	1000	468.4	1800	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 3

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	30.00	0.0	0.0	1500	350.0	2000	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	35.00	0.0	0.0	1700	409.2	2000	2750	3500	10.00	0.0	0.0
180.0	10.00	100.0	40.00	0.0	0.0	1000	468.4	1800	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 4

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	30.00	0.0	0.0	1500	350.0	2100	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	35.00	0.0	0.0	1700	409.2	2100	2750	3500	14.00	0.0	0.0
180.0	10.00	100.0	40.00	0.0	0.0	1000	468.4	2000	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 5

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	40.00	0.0	0.0	1500	350.0	2200	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	45.00	0.0	0.0	1700	409.2	2200	2750	3500	14.00	0.0	0.0
180.0	10.00	100.0	50.00	0.0	0.0	1000	468.4	2200	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 6

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	60.00	0.0	0.0	1500	350.0	2400	2650	3500	8.000	0.0	0.0
90.00	25.00	100.0	60.00	0.0	0.0	1700	409.2	2400	2750	3500	10.00	0.0	0.0
180.0	10.00	100.0	60.00	0.0	0.0	1000	468.4	2200	2700	3500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 7

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	60.00	0.0	0.0	1500	350.0	2600	3550	4500	8.000	0.0	0.0
90.00	25.00	100.0	60.00	0.0	0.0	1700	409.2	2600	3200	4500	14.00	0.0	0.0
180.0	10.00	100.0	60.00	0.0	0.0	1000	468.4	2200	3500	4500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0

Parameter of stimulus 8

	ag	Al	ab	an	ue	f0	f1	f2	f3	f4	ps	dc	ap
0.0	15.00	100.0	60.00	0.0	0.0	1500	350.0	2750	3550	4500	8.000	0.0	0.0
90.00	25.00	100.0	60.00	0.0	0.0	1700	409.2	2750	3200	4500	14.00	0.0	0.0
180.0	10.00	100.0	60.00	0.0	0.0	1000	468.4	2200	3500	4500	8.000	0.0	0.0
190.0	9.818	100.0	60.00	0.0	0.0	1700	475.0	1550	2750	3540	8.000	0.0	0.0
210.0	9.455	100.0	60.00	0.0	0.0	1700	508.8	1531	2760	3540	8.000	0.0	0.0
400.0	6.000	100.0	60.00	0.0	0.0	1700	830.0	1350	2855	3540	8.000	0.0	0.0