INTERRUPTED TONES AND REPEATED SYLLABLES

William Ainsworth*, Georg Meyer* and Jacques Koreman**
*Keele University, UK, **Saarbruecken University, Germany

ABSTRACT

Two harmonically related tones are perceived as a single chord, but if one of them is periodically interrupted they are heard as two separate sound sources. If a /ma/ syllable is repeated the resulting sound consists of a continuous nasal formant with a number of periodically interrupted higher formants. Why is this utterance not perceived as two separate sound sources?

Experiments have been performed to investigate this question. An amplitude modulated tone was added to a continuous tone. Two sources were heard if the stimulus consisted of more than one cycle of amplitude modulation. It was confirmed that repeated /ma/ syllables are perceived as a single voice whereas repeated vowels added to a continuous nasal are perceived as two voices. The signals were subjected to linear prediction analysis then synthesised at the same fundamental frequency. Perception tests showed that both of these stimuli were heard as a single voice.

1. INTRODUCTION

The stream of speech emanating from a single talker is remarkably robust. Although speech consists of sound produced by two acoustic sources, the quasi-periodic glottal pulses of voiced sounds and the random turbulence of fricative sounds, and has a constantly changing spectrum the resulting acoustic stream is received as a single percept.

Two harmonically related tones are perceived as a a single chord, yet if one of them is periodically interrupted the two tones are heard as two separate sound sources [1].

The syllable /ma/ consists of a nasal formant which continues throughout the syllable and a number of higher formants which are weak for the nasal but much stronger for the vowel. So if the syllable is repeated /mamama…/ the resulting sound consists of essentially a continuous nasal formant combined with a number of periodically interrupted higher formants. Why, then, is this utterance not perceived as two separate sound sources? Conversely if a periodically interrupted /a/ vowel is added to a continuous nasal murmur, why does this not sound like a stream of repeated /ma/ syllables?

In order to investigate these questions some experiments have been carried out.

2. TONE EXPERIMENTS

In order to check that the tone experiments of Bregman gave the results reported for tones having frequencies similar to the fomant values in /ma/ syllables a preliminary experiment was performed. A sequence of /ma/ syllables was recorded and analysed (Figure 1) and it was confirmed that there was a band of continuous energy at about 250 Hz and a band of energy at about 1000 Hz modulated at 3-4 Hz. Consequently stimuli were synthesised consisting of a continuous tone at 256 Hz added to an interrupted tone modulated with a 4 Hz square wave. Stimuli were generated varying in duration from 125 ms (half a cycle) to 1250 ms (5 cycles) as described in Table 1.

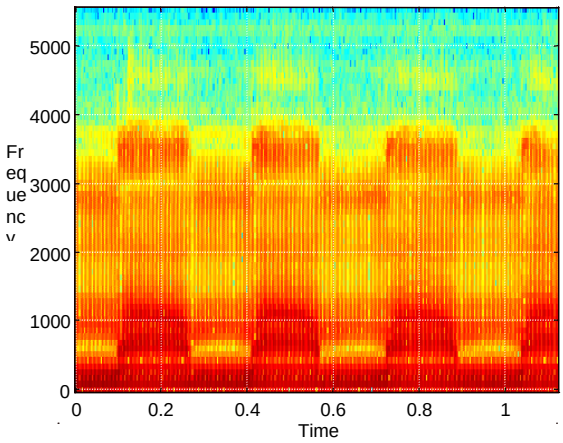

Figure 1. Spectrogram of the repeated syllable /mama…/.

The stimuli were played binaurally over headphones in random order, each 5 times, to listeners who were asked to decide if the sound came from one or two sources. The results are shown in Table 1.

Stimulus	Description	One source (%)	Two sources (%)
1	L, ½ cycle	100	0
2	L+H, ½ cycle	100	0
3	L+H, L, 1 cycle	100	0
4	L, L+H, 1 cycle	40	60
5	L+H, L, 2 cycle	0	100
6	L, L+H, 2 cycle	0	100
7	L+H, L, 5 cycle	0	100
8	L, L+H, 5 cycle	0	100

Table 1. Tone experiments. L is a 256 Hz tone and L+H is a 256 Hz tone plus a 1024 Hz tone. One cycle is 250 ms in duration.

For a half cycle consisting of either a single tone or a dual tone all stimuli were heard as coming from a single source. For a single cycle consisting of a dual tone followed by a single tone the stimuli were also heard as coming from a single source, but for a single cycle consisting of a single tone followed by a dual tone the results were ambiguous. For two or more cycles two sound sources were heard in all cases.

3. SPEECH EXPERIMENTS

The next step was to confirm that a repeated syllable did not sound as though it came from two sources and that adding an interrupted vowel to a nasal murmur did not sound like a sequence of repeated syllables from a single source.

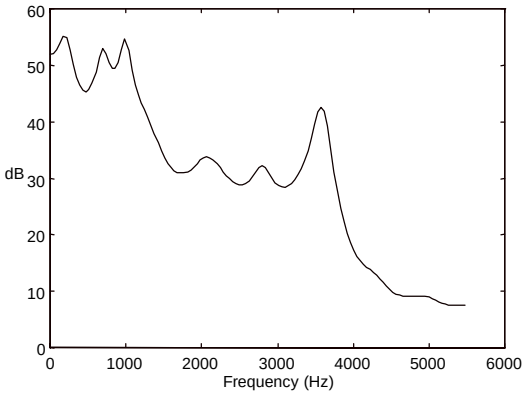

Figure 2. Linear prediction spectrum of the vowel part of the syllable /ma/.

One set of stimuli were derived from a recording of the syllable /ma/ repeated several times. One stimulus (mm11) consisted of a single nasal murmur excised from the sequence, a second (aa11) of a single nasalised vowel, a third (ma11) of a single syllable and a fourth (am11) of a single syllable beginning with a vowel and ending with a nasal, all excised from the sequence. The remaining stimuli (max1 or amx1) consisted of sequences of two or five syllables, where x is the number of syllables.

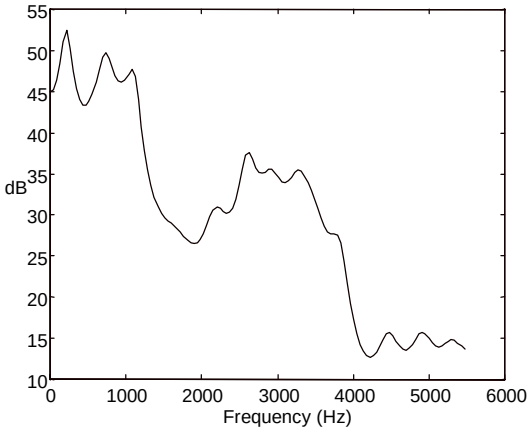

Figure 3. Linear prediction spectrum of the combined signal during the vowel part of the syllable.

The other set of stimuli were formed by adding a continuous nasal /m/ to a sequence of /a/ vowels (combined signal). The nasal sound was attenuated so that the nasal formant was approximately the same amplitude as the first formant of the vowel. Figure 2 shows the linear prediction spectrum of the vowel from a /ma/ syllable and Figure 3 shows the linear prediction spectrum of the vowel part of the combined signal. Nasals (mm12), vowels (aa12) and syllables (max2 or amx2) were excised from the combined signal, where x is again the number of syllables.

The 16 stimuli (half, one, two or five syllables beginning with a nasal or a vowel from one or two sources) were each played to listeners five times in random order binaurally over headphones who were asked if they heard one, two or more voices.

The results are shown in Table 2. Nasals alone were heard as a single voice, but all others were heard as one or two voices according to their means of production.

4. LINEAR PREDICTION EXPERIMENTS

As can be seen from Figures 2 and 3 the vowels in the combined signals have similar spectra to those of the vowels in the syllables, yet they are perceived as emanating from two sources. One possible reason for this is that the nasal murmur was spoken at a higher pitch than the interrupted vowel sequence. Many experiments have shown that concurrent vowels are recognised more accurately if there pitches differ [2, 3, 4, 5] and Brokx *et al*. [6] and Bird and Darwin [7] have shown that words in concurrent sentences are more easily identified if spoken at different pitches. In the present experiments the pitch differences in the combined signals may be preventing the components from fusing into a single percept.

In order to test this possibility new stimuli were synthesised using linear prediction analysis to compute a sequence of vocal tract filters. These filters were excited with sequences of synthesised glottal pulses to generate sounds with the same fundamental frequency throughout the spectrum.

The glottal pulses were produced using a model proposed by Fant [8]. Each pulse consisted of three parts:

$$P1=1/2\ (1-\cos 2\pi ft) \text{ where } f=F0/Fs$$

where F0 is the fundamental frequency and Fs is the sampling frequency

$$P2=k\ (\cos 2\pi ft) - k + 1$$

where k=1.1

$$P3=0$$

The duration of P1 was t1=1/2F0 and that of P2 was t2=1/4F0 and that of P3 was also t3=1/4Fs.

Stimulus	One voice (%)	Two voices (%)
mm11	100	0
aa11	100	0
am11	100	0
ma11	100	0
am21	100	0
ma21	100	0
am51	100	0
ma51	100	0
mm12	80	20
aa12	0	100
am12	0	100
ma12	0	100
am22	0	100
ma22	0	100
am52	0	100
ma52	0	100

Table 2. Speech experiments. The labels represent the stimuli as follows: mm - nasal, aa - vowel, am - vowel-nasal syllable, ma - nasal-vowel syllable, first digit – number of syllables, second digit – number of voices.

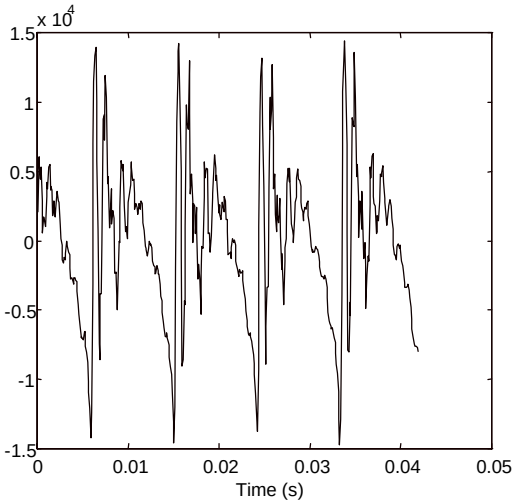

Figure 4. Waveform of the nasalised vowel from a /ma/ syllable.

The speech signal was differentiated then 30 linear prediction coefficients were estimated by the autocorrelation method [9]. The speech signal was synthesised by differentiating the glottal pulse then passing the result through the linear prediction filter. The resulting waveform was similar in shape to the original (Figures 4 and 5). The signal sounded slightly rougher than the original, with better results at lower fundamental frequencies. The stimuli were all synthesised with a fundamental of 88 Hz.

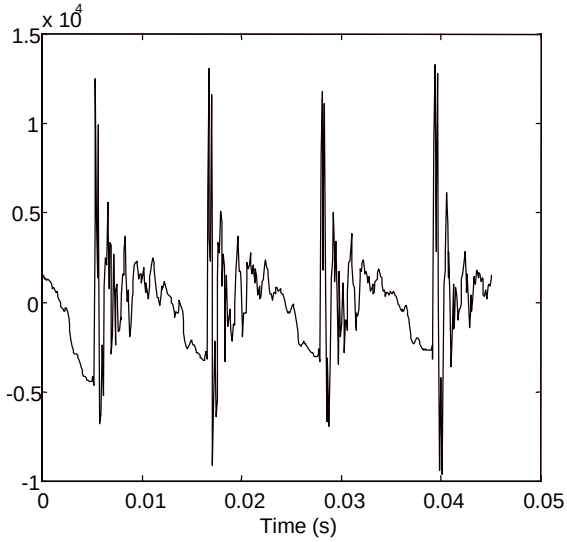

Figure 5. Waveform of the nasalised vowel produced by exciting, with the glottal pulse, the filter obtained by linear prediction analysis of the combined signal formed by adding the nasal to the vowel.

Both the sequence of /ma/ syllables and the combined signal were processed in this way. Sixteen new stimuli, excised from these signals in an analogous way to the last experiment, were produced. These were presented to the same listeners who were again asked to decide if the stimuli sounded as though they were from one, two or more voices. All of the stimuli appeared to come from a single voice.

5. EFFECTS OF PITCH

The results of the previous experiment suggest that continuous /m/ added to interrupted /a/ sounds different from repeated /ma/ because of pitch differences between the two components. In order to confirm this the experiment described in the last section was repeated with the nasal murmur synthesised with a pitch of 120 Hz and the interrupted vowel with a pitch of 88 Hz. The two components were added together after synthesis.

The results are shown in Table 3. The nasal alone and the vowel alone appeared to come from a single voice but all the other stimuli were heard as coming from two voices.

In Section 4 the continuous nasal and interrupted vowel signals were added together then the synthetic stimuli were generated from the combined signal, whereas when the two components had different pitches these were, of necessity, synthesised separately and then added together. In order to test whether this variation in synthesis technique affects the perception of the

stimuli the experiment described in Section 4 was repeated but for half of the stimuli the /m/ and /a/ signals synthesised separately at 88 Hz and then added together whereas for the other half the /m/ was synthesised at 88 Hz and the /a/ at 120 Hz as in the experiment described above.

Stimulus	One voice (%)	Two voices (%)
mm12	100	0
aa12	100	0
am12	20	80
ma12	0	100
am22	0	100
ma22	0	100
am52	0	100
ma52	0	100

Table 3. Pitch experiments. All stimuli were synthesised with 30 linear prediction coefficients and with a fundamental frequency of 120 Hz for the nasal murmur and 88 Hz for the vowels. The labels represent the stimuli as follows: mm - nasal, aa - vowel, am - vowel-nasal syllable, ma - nasal-vowel syllable, first digit – number of syllables, second digit – number of voices.

Stimulus	One voice (%)	Two voices (%)
mm11	100	0
aa11	100	0
am11	80	20
ma11	80	20
am21	60	40
ma21	60	40
am51	80	20
ma51	100	0
mm12	80	20
aa12	100	0
am12	0	100
ma12	20	80
am22	0	100
ma22	0	100
am52	0	100
ma52	0	100

Table 4. Linear prediction experiments. All stimuli were synthesised with 30 linear prediction coefficients and with a fundamental frequency of 88 Hz for the /m/. For stimuli represented by a label in which last digit is 1 the /a/ also had a fundamental of 88 Hz, but for those with a last digit of 2 the fundamental of /a/ was 120 Hz. The labels represent the stimuli as follows: mm - nasal, aa - vowel, am - vowel-nasal syllable, ma - nasal-vowel syllable, first digit – number of syllables.

The results are shown in Table 4. It can be seen that when both the nasal and the vowel were synthesised at the same fundamental the majority of stimuli were heard as coming from one voice but when the nasal and the vowel were synthesised at different fundamentals they were heard as coming from two voices except for the nasal or vowel alone.

CONCLUSIONS

Two harmonically related tones combine to produce a sound which appears to come from a single source, but if one of the tones is periodically interrupted two sound sources are heard. In contrast a sequence of repeated syllables appears to come from a single sound source even though the vowel is periodically interrupted. Adding a nasal murmur to a repeated vowel does not produce a signal which appears to come from a single voice even though its spectrum is similar to that of the syllable. However if the combined signal is resynthesised in such a way that its components all have the same fundamental frequency, it then appears to come from a single sound source.

It is suggested that although non-simultaneous onsets of different frequency components are sufficient to give rise to the percept of multiple sources for simple sounds such as tones this effect is overridden for complex sounds such as speech if the different frequency components have the same fundamental frequency.

REFERENCES

[1] Bregman, A. 1990. *Auditory Scene Analysis*. MIT Press, Cambridge MA.

[2] Sheffers, M. T. M. *Sifting vowels*. Doctoral Dissertation, Groningen University, The Netherlands.

[3] Assmann, P. F. and Summerfield, Q. 1990. Modelling the perception of concurrent vowels: vowels with different fundamental frequencies. *J. Acoust. Soc. Am.*, 88, 680-697.

[4] Culling, J. F. and Darwin, C. J. 1993. Perceptual separation of simultaneous vowels: within and across formant grouping by F0. *J. Acoust. Soc. Am.*, 93, 3454-3467.

[5] Berthommier, F. and Meyer, G. 1995. Source separation by a model of amplitude demodulation. *Proc. Eurospeech'95*,Madrid, 135-13.

[6] Brokx, J. P. L., Nooteboom, G. G. and Cohen, A. 1979. Pitch differences and the intelligibility of speech masked by speech. *IPO Annual Progress Report*, 14, 55-60.

[7] Bird, J. and Darwin, C. J. 1998. Effects of difference in fundamental frequency in separating two sentences. In *Psychophysical and Physiological advances in Hearing* (A. R. Palmer, A. Rees, A. Q. Summerfield and R. Meddis, Eds.), Whurr Publishers Ltd., London, 263-269.

[8] Fant, G. 1979. Glottal source and excitation analysis. *STL-QPSR*,2-3, 1-19.

[9] Jackson, L. *Digital Filters and Signal Processing*, 255-257.