AMPLITUDE MODULATION OF FRICATION NOISE BY VOICING SATURATES

Jonathan Pincas and Philip J.B. Jackson [j.pincas, p.jackson]@surrey.ac.uk
Centre for Vision, Speech and Signal Processing, University of Surrey, GU2 7XH, UK

2CP3-13

1. ABSTRACT

• Voiced fricatives: two major speech sources, voicing and frication simultaneously.

• Characteristics of each source are influenced by the other: mutual interaction effects.

• One interaction effect is amplitude modulation of the frication source near the constriction by the periodic source at the glottis (fundamental frequency, f_0).

• Modulation index (or depth), m, was measured for a corpus of English sustained fricatives using a novel technique involving high-pass filtering, spectral analysis in the modulation domain and jitter/pitch glide compensation.

• At low voicing strengths, m rises linearly as voicing strength increases.

• At a voicing strength of 66 dB SPL, m saturates and ceases to rise.

• The value of m at saturation depends upon speaker and fricative place of articulation, with sibilants (in particular [z]) displaying higher levels of modulation.

2. MODULATION OF FRICATION: BACKGROUND

Fant (1960):	Noted that source interaction occurred as "periodic and synchronous" modulation of the frication source by phonation.
Klatt (1980):	Incorporated modulation for voiced fricatives into speech synthesiser.
Jackson and Shadle (2000):	Measured modulation depths and phases for fricatives. Suggested a possible mechanism for modulation.
Pincas and Jackson (2004):	Investigated modulation indices for a corpus of English fluent-speech fricatives.

3. WHAT IS MODULATION OF FRICATION NOISE?

• A modulated noise signal, $x(n)$, can be expressed as a modulating signal, $a(n)$, multiplied by a noise source, $w(n)$, acting as the carrier.

• In the case of sinusoidally amplitude-modulated (SAM) noise, $a(n)$ is a pure tone.

• $a(n)$ need not be (and in real speech is unlikely to be) perfectly sinusoidal. For any periodic $a(n)$, $x(n)$ is described by Equation 1.

• SAM noise is pulsed at a specified frequency, f_0. In speech, f_0 corresponds to the voicing fundamental frequency. Depth of modulation, m, is the level of RMS fluctuation compared to the d.c. level expressed as a fraction (see waveforms of noise modulated with different m in Figure 1.)

• In voiced fricatives, modulation often goes unnoticed as it is hard to detect in the waveform which also contains voicing. Pulsed noise can be brought to the foreground by high-pass (HP) filtering to remove the voicing component and by taking the magnitude of the signal. Figure 2 illustrates this for a token of [z]. Also notice in Figure 3 the vertical striations in the spectrogram as the amplitude of the noise throughout the spectrum periodically rises and falls.

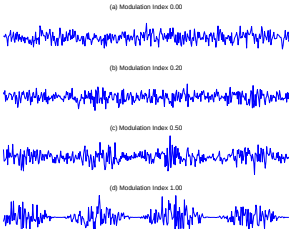

Figure 1: Broadband noise modulated with (from top) m={0, 0.2, 0.5, 1} (f_0= 150 Hz, f_s=8 Hz).

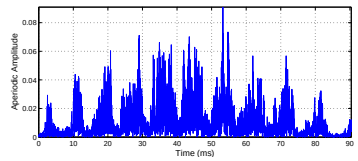

Figure 2: Instantaneous magnitude signal for a 90 ms token of [z] spoken by JP. f_{HP} =3 kHz to isolate frication noise.

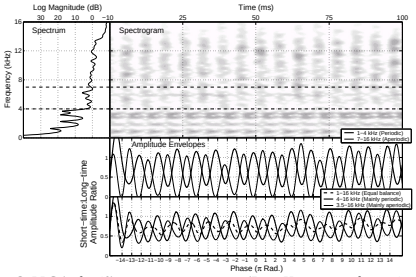

Figure 3: LPC (order 40) spectrum ,spectrogram (5 ms, Hanning window, ×4 zero-padded, fiexed gray-scale) and amplitude envelopes (magnitude signal, low-pass filtered at 200 Hz) for 100 ms section of sustained voiced fricative [v] ($f_0 \approx 153$ Hz, bandwidth = 16 kHz). Individual amplitude envelopes are for different frequency bands, f_{BP}. Upper: f_{BP} = 1–4 kHz (thick line, periodic energy only) and f_{BP} = 7–16 kHz (thin line, aperiodic energy only); dashed horizontal lines on spectrogram identify these frequency regions. Lower: f_{BP} = 1–16 kHz (thick line, mainly periodic), f_{BP} = 4–16 kHz (thin line, mainly aperiodic) and f_{BP} = 3.5–16 kHz (dashed line, balanced mix of periodic and aperiodic).

8. CONCLUSION

• Developed a process for estimating modulation index from speech recordings.

• Voiced fricatives in speech are *always* modulated. The depth of modulation depends to some extent on fricative place of articulation and speaker characteristics.

• The results accord well with measurements obtained for fluent-speech fricatives (e.g., Pincas (2004)), although sustained fricatives tend to display higher levels of modulation.

• Although it is likely to depend on some form of regularisation of the turbulence formation process through acoustic wave interaction, the exact details of the modulation procedure are still unclear. Aerodynamic simulations using vocal tract models are proposed.

• The perception of modulation in frication noise also requires investigation. The psychoacoustic literature suggests that the levels of modulation observed here should be perceptible, although it is not clear whether it could function as any form of linguistic cue. The perception of frication modulation in speech is currently being investigated

References

Crow, S. and F. Champagne (1971). Orderly structure in jet turbulence. *Journal of Fluid Mechanics 48*(3), 547–591.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague, Netherlands: Mouton.

Jackson, P. J. B. and C. H. Shadle (2000, October). Frication noise modulated by voicing, as revealed by pitch-scaled decomposition. *J. Acoust. Soc. Am. 108*(4), 1421–1434.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *J. Acoust. Soc. Am. 67*(3), 971–995.

Pincas, J. (2004). The interaction of voicing and frication sources in speech: An acoustic study. Master's thesis, University of Surrey.

Pincas, J. and P. J. Jackson (2004, June). Acoustic correlates of voicing-frication interaction in fricatives. In *Proceedings of the 'From Sound to Sense' Conference, MIT, Cambridge, MA, USA.*

Web: http://www.ee.surrey.ac.uk/Personal/[j.pincas,p.jackson]/

BASIS OF THE ESTIMATION PROCEDURE FOR RETRIEVAL OF AMPLITUDE MODULATION INDEX, m, FROM FRICATION NOISE

For any periodic modulating signal:

$$x(n) = w(n)\left(1 + \sum_{h=1}^{H} m_h \cos\left(h\omega_0 n + \phi_h\right)\right), \qquad (1)$$

where $w(n)$ is the noise signal, $h \in \{1..H\}$ are the harmonics, $\omega_0 = 2\pi f_0/f_S$ is the normalised angular frequency for the fundamental modulation frequency f_0, f_s is the sampling frequency, m_h is the modulation index at hf_0, and ϕ_h is the phase shift.

Take the Fourier Transform of the magnitude signal, $y(n) = |x(n)|$:

$$Y(k) = \mathcal{F}\{|w(n)|\} \otimes \left(\delta(k) + \sum_{h=1}^{H} \frac{m_h}{2}\left(\delta(k \pm hk_0)\, e^{\pm j\phi_h}\right)\right), \qquad (2)$$

where $\otimes$ denotes convolution, $\delta(\cdot)$ is the Dirac delta function, and $k_0 = Nf_0/f_S$ is the frequency bin that contains f_0.

For m_1, compare coefficients around f_0 to those around d.c.:

$$\hat{m}_1 = 2\left(\frac{\sum_{\hat{k}_0} Y(k)^* Y(k) - \theta^2}{\sum_{\bar{0}} Y(k)^* Y(k) - \theta^2}\right)^{\frac{1}{2}}, \qquad (3)$$

where * denotes complex conjugate, θ^2 is the noise power, and $\hat{k}_0$ and $\bar{0}$ cover the spectral peaks around k_0 and 0 respectively.

Evaluation of accuracy:
Simulations on modulated white noise, including effects of pitch glide and jitter, indicated 2-σ error ranges of ± 0.04 (100 ms window) and ± 0.08 (25 ms window) on estimates of the modulation index.

4. SUSTAINED FRICATIVE CORPUS

The Corpus: Fricative tokens, F=/v,ð,z,ʒ/ (i.e., /v, dh, z, zh/), uttered in isolation, 16 speakers (12 M, 4 F), two types of utterance at three pitch settings, f_0={125,150,175 Hz}: uninterrupted fricative with varying effort and three separate sustained fricative bursts with stepped effort.

5. METHOD: APPLICATION OF ESTIMATION PROCEDURE TO VOICED FRICATIVES

• **Variable pitch:** pitch is rarely constant through the analysis window leading to spread of energy around f_0 in spectral and modulation domains. SAM noise where f_0 is constant throughout the window has components in the modulation spectrum at d.c. and f_0 only, and all of the modulating signal's energy is at f_0; in this case, m is simply calculated as $2\,|Y(k_0)|\,/\,|Y(0)|$. To tackle spread of energy around f_0 we integrate over the peak rather than taking the value from a single frequency bin.

• **Presence of periodic energy:** before estimation of m, frication noise must be isolated from periodic energy (i.e., f_0 and harmonics/formants). Formants with periodic energy are also modulated at f_0, so their presence serves to boost or attenuate modulation depth of noise depending on the phase relationship between frication bursts and glottal vibration. This is illustrated in Figure 3: amplitude envelopes are shown for various spectral regions of the signal. The fricative and voicing envelopes are out of phase (upper plot) causing a less modulated envelope if not separated (lower plot). HP filtering is used with a cuton frequency f_{HP} high enough to exclude the periodic energy.

• **Different spectral regions modulated to different degrees:** frication noise is not always evenly modulated across the spectral range, with higher spectral regions exhibiting greater modulation. HP filtering with higher cut-on frequencies, whilst effectively eliminating periodic energy, will emphasise spectral regions that are more heavily modulated, leading to over-estimation of m.

7. DISCUSSION

• **Shape of $v_1 m_1$ function:** rises linearly up to 0.04-0.06 Pa SPL (66-70 dB SPL) where saturation occurs. Little variation in saturation point across places of articulation and speakers.
• **Place of articulation:** sibilant/non-sibilant split, [z] most strongly modulated. Generally consistent across speakers.
• **Effect of varying f_{HP}:** m at saturation rises as f_{HP} increases. At low f_{HP} values (0.7, 1.4 kHz), this is presumably due to cutting out periodic energy (see Figure 3); at higher f_{HP}'s (8.4, 11.5 kHz) it is probably due to isolation of strongly modulated frication noise. Middle f_{HP} values (2.7, 4.5 kHz) give a more realistic overall picture.
• **Harmonic structure and modulation mechanism:** m_1 (modulation component at f_0) is dominant but harmonic structure in the modulation domain emerges with m_2 saturating around m=0.15. Jackson and Shadle (2000) have suggested that modulation results from forcing (regularisation) of turbulence generation at source by the low frequency pressure wave. Raises the issue of the connection between the modulating signal and the physical wave that is responsible for the forcing mechanism. Harmonics in the modulation spectrum may not result from a harmonics in the voicing (Crow and Champagne (1971) showed that *sinusoidal* forcing can produce modulation with a harmonic). Similarly, harmonic structure in the audio domain may not be replicated in the modulation domain (see Figure 8).

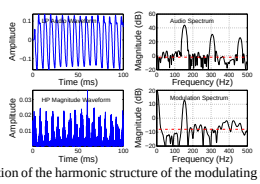

Figure 8: Illustration of the harmonic structure of the modulating signal and voicing component for portion of fricative [z] ($f_0 \approx 150$ Hz). Top left: audio waveform low-pass filtered at 1 kHz. Top right: audio spectrum up to 500 Hz. Bottom left: magnitude waveform high-pass filtered at 9 kHz. Bottom right: its modulation spectrum. Dashed lines indicate noise floor.

Cut-on frequencies $f_{HP} \in \{0.7, 1.4, 2.7, 4.5, 8.4$ and 11.5 kHz$\}$

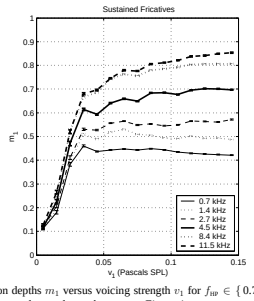

Figure 6: Modulation depths m_1 versus voicing strength v_1 for $f_{HP} \in \{0.7, 1.4, 2.7, 4.5, 8.4$ and 11.5 kHz$\}$. Binning procedure and error bars as per Figure 4.

Modulation at the harmonics: m_1, m_2 and m_3

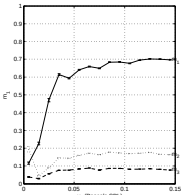

Figure 7: Modulation depths at the fundamental frequency m_1, second harmonic m_2 and third harmonic m_3 versus voicing strength v_1. Means from all tokens at f_{HP}=4.5 kHz. Binning procedure and error bars as per Figure 4.

6.RESULTS

Voicing strength, v_1, plotted against modulation index, m_1, for a variety of conditions:

Places of articulation

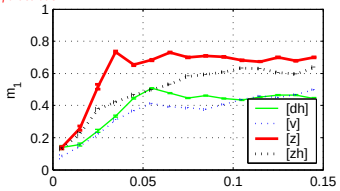

Figure 4: Modulation depths m_1 versus voicing strength v_1. Data for each fricative are means across all tokens.Data are means of values falling within ± 0.005 Pa bins from all tokens. Error bars show standard error (SE).

Places of articulation for individual speakers

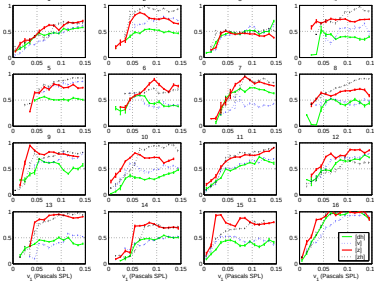

Figure 6: Modulation depths m_1 versus voicing strength v_1 for individual speakers for sustained fricatives /ð,v,z,ʒ/ at f_{HP}=4.5 kHz. Data for each fricative are means across all tokens. Binning procedure and error bars as per Figure 4.