ALTURA DE TONOS CON VIBRATO: UN MODELO BASADO EN LA DESCOMPOSICIÓN EN FRECUENCIAS INSTANTÁNEAS

Bruno A. Mesz, Manuel C. Eguía.

LABORATORIO DE ACÚSTICA Y PERCEPCIÓN SONORA. UNIVERSIDAD NACIONAL DE QUILMES

Resumen

Estudiamos el vibrato como la manifestación más significativa de un tono no estacionario que suscita la percepción de una única altura global. Algunos resultados recientes concernientes a tonos con vibratos no simétricos sugieren que la altura percibida podría depender de algún mecanismo sensible a la estabilidad en frecuencia (Gockel et al. 2001). Mostramos que una representación tiempo-frecuencia propuesta recientemente (Gardner y Magnasco 2006) podría ser el marco adecuado para explicar este hipotético mecanismo. Proponemos un modelo uniparamétrico dentro de este marco que puede dar cuenta de datos experimentales sobre vibrato previamente publicados así como de nuevos datos obtenidos en experimentos psicofísicos en nuestro laboratorio.

Introducción

La amplia mayoría de los estudios en percepción de altura se concentra en estímulos estacionarios, especialmente tonos complejos, formados por múltiples frecuencias fijas. Sin embargo, en los sonidos naturales y en música, los componentes en frecuencia usualmente varían en el tiempo, a menudo en forma coherente. Los tonos armónicos complejos, por ejemplo, tienen una frecuencia fundamental (F0) bien definida, que está fuertemente correlacionada con la altura percibida para tonos estacionarios. Cuando estos tonos se producen en un contexto natural, la F0 nunca es constante y la asignación de altura es menos clara. Sin embargo, bajo ciertas circunstancias, podemos percibir una sola altura global (llamada altura principal) de un tono cuya F0 fluctúa. El ejemplo más usual es el vibrato musical.

El vibrato es una modulación casi periódica de los componentes en frecuencia de una nota, en un rango que excede a veces el semitono. En contraste con la opinión corriente de que el vibrato puede usarse para disimular una pobre entonación, estudios psicofísicos han mostrado que la altura puede extraerse con un limen menor que un décimo de semitono (van Besouw et al. 2008). Cuál sea esa altura ha sido no obstante materia de debate.

Shonle y Horan (1980) obtienen que la media geométrica de las F0 es una buena aproximación a la altura percibida. Iwamiya et al. (1984) presentan una conclusión similar. Sin embargo, d'Alessandro and Castellengo (1994) encuentran que la parte final de la modulación es la más importante para la percepción global. Shonle y Horan (1980) reportan también que para modulaciones asimétricas la altura principal difiere de la media geométrica y sugieren un promedio temporal diferente de las frecuencias presentes en el estímulo.Gockel et al. (2001) confirman estos corrimientos de altura para sinusoides moduladas asimétricamente y proponen un model donde el promedio de frecuencias está dado por un peso que decrece al crecer la tasa de modulación. En otras palabras, si la modulación tiene porciones de variación lenta y rápida de F0, las partes de variación lenta son las que contribuyen más significativamente a la altura principal. Los autores llaman a este promedio "sensible a estabilidad" ("*stability-sensitive weighting*")

En este trabajo mostramos que este mecanismo de promedio sensible a estabilidad puede ser derivado de una representación no lineal en análisis tiempo-frecuencia propuesta recientemente por Gardner and Magnasco (2006, 2005). Esta representación se conoce como método de reasignación y proporciona una representación precisa de sonidos rápidamente variables, en particular del vibrato. De hecho, la parte esencial del método es la noción de "consenso" que calcula la consistencia de reasignaciones de tiempo y frecuencia en distintos "canales" (explicaremos esto más en detalle en la sección dedicada al modelo). Nuestra hipótesis de trabajo es que el mecanismo sensible a estabilidad descrito en Gockel et al. (2001) es una consecuencia directa del criterio de consenso definido en Gardner y Magnasco (2006).

A efectos de poner a prueba nuestra hipótesis formulamos un modelo simple basado en el método de reasignación y en una medida de consenso, y realizamos un experimento psicofísico que explora paramétricamente la transición entre una modulación simétrica y otra altamente asimétrica.

Experimento

Nuestros estímulos consistieron en sinusoides moduladas en frecuencia. Utilizamos perfiles de modulación trapezoidales, consistiendo cada ciclo de una porción plana (frecuencia constante) y dos segmentos lineales formando un pico. El porcentaje del ciclo ocupado por la porción plana (p) era variable. Los picos podían apuntar hacia las altas frecuencias (lo que designaremos como perfiles **u**) o hacia las bajas frecuencias (perfiles **n**). Para p=0 la forma de la modulación es triangular y no hay distinción entre perfiles **u** y **n**.

Para todos los estímulos empleamos cuatro ciclos de modulación con una media geométrica de 1000 Hz, una tasa de modulación de 10 Hz y una profundidad de modulación pico a pico de 150 cents. La fase inicial de la modulación se varió al azar para compensar los efectos de una posible mayor importancia del comienzo o fin del vibrato.

Cuatro sujetos participaron en el experimento, todos de audición normal con un grado variable de entrenamiento musical y edades variando entre 25 y 38 años. Dividimos las sesiones experimentales en siete bloques, uno para cada tipo de modulación: p=0 (simétrica) y p=25, 50, 75 para perfiles **u** y **n**. En cada bloque se obtuvieron cuatro ajustes de altura entre el tono modulado y una sinusoide. Para obtener los ajustes se empleó un método de elección forzada de dos alternativas y una escalera de paso adaptativo.

Resultados

En la figura 1 se muestran los resultados, promediados respecto de los sujetos. Los valores que se muestran son las frecuencias de ajuste promedio junto con sus desviaciones standard, para los siete estímulos presentados. En el caso de la modulación simétrica (p=0) la frecuencia ajustada promedio está próxima a la media geométrica, como era lo esperado. Para modulaciones no simétricas obtuvimos corrimientos de altura consistentes con los resultados previos: las frecuencias medias ajustadas se desviaban en dirección de la porción plana y más estable de la modulación.

Modelo

Pasamos ahora a la formulación de un modelo de altura principal basado en el método de reasignación. Este modelo se basa en la información de fase obtenida de la señal acústica (una información bien preservada en el sistema auditivo periférico). El modelo tiene tres etapas: a) extracción de frecuencias instantáneas, b) comparación de las frecuencias entre distintos canales (cálculo de consenso) y c) un promedio por consenso y amplitud.

Comenzamos con la Tranformada de Fourier a tiempo corto (Short-Time Fourier Transform o STFT) de la señal acústica x(t), calculada en un conjunto discreto de frecuencias a las que nos referiremos como canales. La STFT puede ser reescrita en términos de su amplitud X y fase Φ

$$STFT(t,\omega) = X(t,\omega)e^{\Phi(t,\omega)} = \int x(t+\tau)h(-\tau)e^{-i\omega t}d\tau \qquad (1)$$

Definimos las frecuencias instantáneas canalizadas, (channelized instantaneous frequencies o CIF) como la derivada temporal de la fase de la STFT :

$$CIF(t,\omega) = \frac{\partial}{\partial t}\Phi(t,\omega) \qquad (2)$$

Estimamos luego la frecuencia instantánea de la señal como el promedio de las CIF respecto de todos los canales de análisis, pesada por la energía por canal (calculada como el cuadrado de X).

$$FI(t) = \frac{\sum_{\omega} CIF(t,\omega) X(t,\omega)^2}{\sum_{\omega} X(t,\omega)^2} \qquad (3)$$

El módulo de la MPD mide la tasa de cambio de las CIF a través de los canales. Un módulo grande (respectivamente, pequeño) de la MPD puede ser visto como indicativo de un alto (respectivamente, bajo) consenso. Estas consideraciones nos conducen a considerar una función de peso perceptual W1 que toma un valor máximo cuando existe un máximo consenso (identidad de las CIF de distintos canales) y decrece exponencialmente al disminuir el consenso. Esta función de peso se promedia también proporcionalmente a la energía por canal:

$$MPD(t,\omega) = \frac{\partial}{\partial \omega}\left(\frac{\partial}{\partial t}\Phi(t,\omega)\right) = \frac{\partial}{\partial \omega} CIF(t,\omega) \qquad (4)$$

$$W1(t;\gamma) = \frac{\sum_{\omega} X(t,\omega)^2 \exp\left(-|MPD(t,\omega)|/\gamma\right)}{\sum_{\omega} X(t,\omega)^2} \qquad (5)$$

Utilizamos γ como nuestro parámetro de ajuste a los datos experimentales. Notemos que un valor grande de γ da un peso que tiende a ser independiente del consenso, mientras que un valor pequeño de γ privilegia los puntos de alto consenso.

Para incluir vibratos que presentan también modulación en amplitud, añadimos un segundo peso W2 igual a la raíz cuadrada de la suma de la energía en todos los canales:

$$W2(t) = \sqrt{\sum_{\omega} X(t,\omega)^2} \qquad (6)$$

Finalmente, la altura principal PP se calcula promediando en el tiempo:

$$PP(\gamma) = \frac{\sum_{t} FI(t) W1(t;\gamma) W2(t)}{\sum_{t} W1(t;\gamma) W2(t)} \qquad (7)$$

Examinamos la aplicabilidad del modelo a datos previos sobre percepción de altura principal en vibrato (Gockel et al. 2001, Shonle y Horan 1980, Iwamiya et al. 1984). En todos los casos pudimos ajustar los datos dentro de las desviaciones standard reportadas usando un solo valor de γ=0.04 (verTabla 1). Hay en realidad un rango amplio de valores de γ que ajustan razonablemente. Las predicciones para nuestro experimento con el mismo valor de γ=0.04 se muestran en la Figura 1 como líneas punteadas y observamos nuevamente una buena aproximación.

Conclusiones

Presentamos un modelo para la percepción de altura principal en tonos con vibrato que predice resultados anteriores y nuetras nuevas observaciones. El modelo propuesto tiene cierta plausibilidad biológica ya que los cálculos de frecuencia en cada canal se basan en información de fase, por lo que en principio podrían ser implementados usando el intervalo temporal entre potenciales de acción en el nervio auditivo. Ha habido por mucho tiempo debate sobre si la percepción de altura está relacionada con la codificación de lugar y tasa de espigas en el eje tonotópico o con la codificación temporal. Nuestro modelo integra la organización tonotópica (relacionada con la amplitud de los canales) , codificación temporal (via las CIF), y un chequeo

cruzado entre canales de ambos tipos de información (consenso), para predecir una única altura percibida.

Agradecemos a M. Magnasco por discusiones útiles.

Referencias

Gockel, H., B. C. J. Moore & R. P. Carlyon (2001). Influence of rate of change of frequency on the overall pitch of frequency-modulated tones. J. Acous. Soc. Am. **109: 701-712.**

Gardner, T. J. & M. O. Magnasco (2006). Sparse time-frequency representations. Proc. Natl. Acad. Sci. **103: 6094-6099.**

van Besouw, R. M., J. S. Brereton & D. M. Howard. (2008). Range of tuning for tones with and without vibrato, Music Perception **26: 145-156**.

Shonle, J., & K. Horan.(1980). The pitch of vibrato tones. J. Acous. Soc. Am. **67: 246-252**.

Iwamiya, S., K. Kosugi & O. Kitamura (1984). Perceived principal pitch of FM-AM tones when the phase difference is in-phase and anti-phase. J. Acoust. Soc. Jpn. **5: 59-69**.

D'Alessandro, C. & M Castellengo (1994). The pitch of short-duration vibrato tones. J. Acous. Soc. Am. **95: 1617-1630**.

Gardner, T. J. & M. O. Magnasco (2005). Instantaneous frequency decomposition : an application to spectrally sparse sounds with fast frequency modulations. J. Acous. Soc. Am. **117: 2896-2903**.

Tabla 1. Resultados previamente reportados de altura principal en vibratos no simétricos (Gockel et al. 2001, Shonle and Horan (1980). La frecuencia central (CF) es la media geométrica de la modulación. La predicción de nuestro modelo se hizo con el mismo valor del parámetro (g=0.04) que ajusta nuestros resultados experimentales. También ajustamos los resultados de Iwamiya et al. (1984) (datos no incluidos).

Modulation shape	F (Hz)	M depth (cents)	M degree	M rate (Hz)	Starting Phase (radians)	ean (Hz)	td (Hz)	rediction (Hz)	Reference
Trapezoidal U	368	100	0	6	0	362	14	362	Shonle and Horan [9]
Trapezoidal ∩	524					546	16	538	
Non-symmetrical U	1000	150	0	0	Pi	989.6	2.6	989.8	Gockel et al. [1]
					0	990.0	2.2	988.7	
Non-symmetrical ∩					Pi	1009.7	2.1	1010.7	
					0	1011.2	2.0	1011.7	

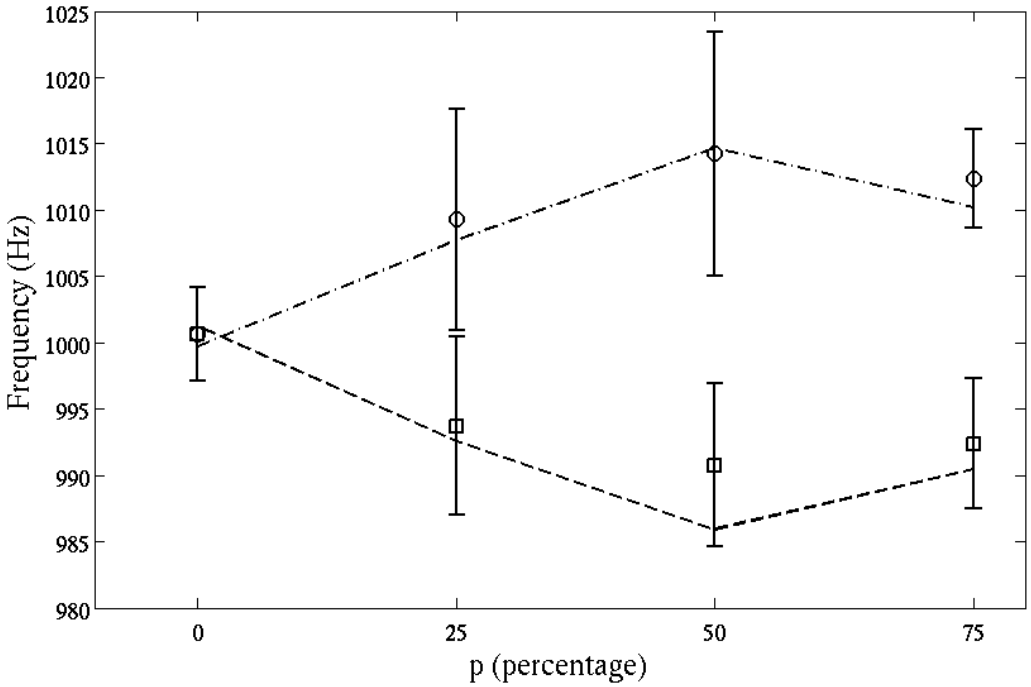

Figura 1. Corrimientos de altura observados para la altura principal de las sinusoides moduladas descritas en el texto. Para los perfiles **u** profiles (cuadrados) los cuatro sujetos reportaron alturas principales por debajo de la media geométrica de 1 kHz. Los tonos con perfiles **n** (círculos) evocaron alturas mayores que la media. La predicción de nuestro modelo (ver ecuación 7) para un valor del parámetro g=0.04 y para perfiles **u** profiles (perfiles **n**) aparece dibujada en líneas discontinuas (puntos y líneas discontinuas).