高速度カメラを用いた音収録

Acoustical Measurement with High-Speed Camera

5111E001-1　阿久津　真理子　　指導教員　及川　靖広　教授
AKUTSU Mariko　　Prof. OIKAWA Yasuhiro

概要：本研究では高速度カメラで撮影した話者映像からの発話情報取得を試みた。マイクロホンを用いない音収録の手法として高速度カメラを用いた研究が行われてきた。高速度カメラでの収録は，カメラの画角内であれば収録点から離れた点から多点同時収録できるので，音場に影響を与えない，雑音に強く音源分離も容易であるという利点がある。一方，音声コミュニケーションでは，現在広く研究が行われている音声認識やコンピュータでの自動読唇の認識結果として返されるような文字情報に加えて声質やイントネーションも相手の意図をくみ取る上での重要な情報となる。そこで，高速度カメラを用いて映像のみからの声質等の情報を含めた音声の再現を目指し，コミュニケーション手段としての可能性を検討する。

キーワード：高速度カメラ，発話情報，頸部振動，音源分離，イントネーション
Keywords: high-speed camera, voice information, cervical part vibration, sound source separation, intonation

１．まえがき

通常音の収録にはマイクロホンが用いられることが多いが、収録音場に直接設置する必要があるので、マイクロホン自体が測定音場に影響を与える、広域を収録する際に多数のマイクロホンが必要となるという問題があった。そこでマイクロホンを用いない音収録手法が研究されてきた[1][2]。そのひとつに高速度カメラを用いた音の収録がある。高速度カメラで被写体の振動を記録し，被写体の動きから音の振動を取り出すので，画角内の多点を同時記録可能であり関心領域を自由に設定できる。このことから，音源分離が容易，雑音に強いといった特徴があり，音源や，音場情報の取得に関する研究が行われてきた[3][4][5]。

本研究では人が発話している際の頸部の高速度映像から発話情報取得を行った。高速度カメラで音の収録を行うことで，音声認識や自動読唇の結果として返される文字情報に加えて，声質やイントネーション等の個人性や話者の気分を示すような情報が取得可能であると考えられる。

２．高速度カメラを用いた音収録

２.１．高速度映像からの音収録

本研究では高速度映像から音による振動を含んだ被写体の動きを取得したが，今回は画素値の重心位置変化を用いて振動を追跡した。カメラで撮影した映像を見ると，対象物の動きに伴って各画素の画素値が変化していることが分かるので画素値の重心位置の変化を対象物の位置変化と見なすことができる。

撮影したグレースケール画像中のある領域の大きさを H px×W px とし領域内の画素値の合計を M とすると，重心座標(Gi,Gj)は画素(i,j)の画素値$g[i][j]$を使って

$$\begin{bmatrix} G_i \\ G_j \end{bmatrix} = \frac{1}{M} \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} g[i][j] \begin{bmatrix} i \\ j \end{bmatrix} \qquad (1)$$

と計算される。このように画像フレーム毎に物体の画像重心位置を検出し，この時間変化を物体の振動とする。

２.２．高速度カメラを用いた音源分離

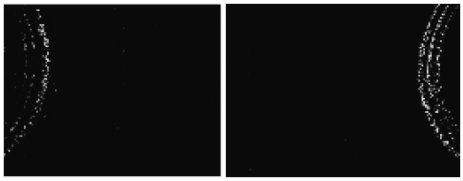
高速度カメラで振動記録する場合，画角内の情報を同時に記録できるのでそれぞれの位置の振動を取り出すことが可能である。実際に２台のスピーカを異なる音で同時に駆動したものを高速度カメラで撮影した。2 台のスピーカを左右に並べ，向かって左のスピーカを 530 Hz，右のスピーカを 820 Hz で駆動した。撮影した画像を図―1 (a) に示す。使用した高速度カメラは MEMRECAM HX-3 (nac image technology)，6,000 fps で撮影した。10 px×10 px 範囲毎の画素値重心位置の時間変化を求め，その周波数を得た。画像中の 500-550 Hz で振動している領域を表示すると図―1 (b),

800-850 Hz で振動している領域を表示すると図—1 (c)となり，それぞれスピーカの位置に表示されていることが分かる。

３．頸部高速度映像からの発話情報取得

20 代の男子学生が，/meguro/と発声している様子を高速度カメラで撮影した。また，参照用として，高速度カメラでの撮影と同時に頸部の振動をレーザドップラ振動計，音声をマイクロホンで収録した。レーザドップラ振動計での頸部振動収録では測定精度を上げる目的で被験者の頸部に反射テープを貼り，そこにレーザを照射した。

(a)撮影画像

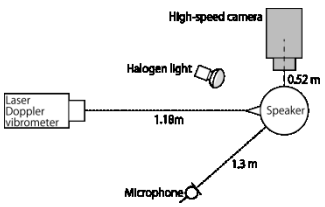


(b)500-550 Hz 表示　　(c)800-850 Hz 表示

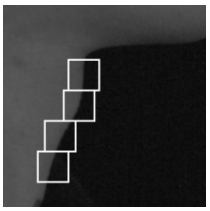
図—1　2 台のスピーカの撮影と解析結果

表—1 実験詳細

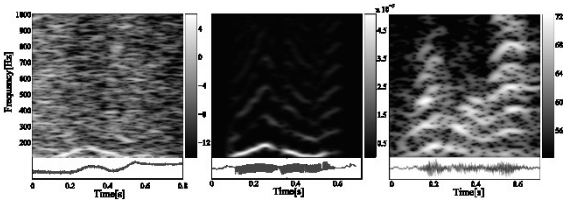
High-speed camera	MEMRECAM HX-3 (nac image technology)
Frame rate	8,000 fps
Laser Doppler vibrometer	OFV-505 (Polytec)



図—2 実験配置



図—3 解析範囲



(a)高速度カメラ (b)レーザドップラ振動計 (c)音声

図—4 /meguro/ スペクトログラム

この際の実験条件を表—1 に，配置を図—2 に示す。撮影した頸部と解析範囲は図—3 である。高速度カメラで撮影した頸部の振動，レーザドップラ振動計で記録した頸部振動，マイクロホンで収録した音のスペクトログラムを図—4 (a)，図—4 (b)，図—4 (c)に示す。それぞれの波形をスペクトログラム下に示す。高速度カメラのデータは 4 ヶ所の振動を周波数領域で平均し，表示したものである。これらの図より，イントネーションである発話中の周波数の揺れが高速度カメラの振動に再現されていることが確認できた。

４．むすび

マイクロホンを用いない音の収録方法として提案されている高速度カメラを用いた，発話情報取得に関する考察を行った。本稿では高速度カメラでの収録の利点を生かして音源分離を行い，また話者の頸部の振動を高速度カメラで撮影し，解析を行った。頸部からの発話情報取得の際には，参照用にレーザドップラ振動計で記録した頸部振動とマイクロホンで記録した音声と比較した。その結果高速度カメラの映像から取得した振動には，音声に含まれている周波数成分と共通した成分やイントネーションの揺れが含まれていることが分かり，高速度映像からの音声の再現の可能性が示された。

参考文献

[1] Y. Oikawa *et al*, “Sound Field Measurement Based on Reconstruction from Laser Projections,” Proc. ICASSP, IV, pp.661-664, 2005.

[2] 園田義人, “レーザ光による音波・超音波検出と光マイクロホンの基礎研究,” 音講論(秋), pp.599-600, 1995.

[3] 酒井寿理他, ”高速度カメラを用いた 3 次元音場の記録と解析,” 音講論(秋), pp.1493-1494, 2008.

[4] 武岡成人他, “高速度カメラを用いた PIV 法による粒子速度分布の測定,” 日本音響学会アコースティックイメージング研究会資料, AI2010-3-03, 2010.10.

[5] M. Akutsu *et al*, “Extraction of sound field information from flowing dust captured with high-speed camera,” Proc. ICASSP, pp.545-548, 2012.3.