

目が合う自然な会話を可能とする 対面型ビデオ通話システム

移動機開発部 きむら しんじ 木村 真治 おおせき えりこ 大関 江利子

情報通信技術の進歩とともに普及が進むビデオ通話におけるユーザ体験の向上には、対面感の実現が肝要である。そのためには画質や画面サイズ向上などの多くの技術が必要だが、実際のサービス提供のためには、リーズナブルなコストで、実用上十分な対面感を実現するシステムが求められる。そこで、ドコモはビデオ通話における各パラメータの対面感への寄与を評価し、特に重要な寄与を示したアイコンタクトに着目し、これを実現する正面撮影技術をベースとした対面型のビデオ通話システムを開発した。本稿では、その詳細を解説する。

1. まえがき

遠隔地間でのビデオ通話は、カメラやディスプレイなどのデバイスの進化、ネットワークの高速化、スマートフォンやPCなどのツールの普及に伴い、友人間などでのカジュアルなコミュニケーション用途や、ビジネスにおけるミーティングなどで一般的に用いられるようになった。このような映像コミュニケーションでは、遠隔地間で「通話相手があたかも同じ場において会話をしている感覚（以下、対面感）」を実現することが、ユーザ体験向上のための大きな目的の1つであるが、実際の対面での会話や、会議に完全にとって代わるものとして利用されてい

るとは言い難いのが現状である。

その要因として、心理的・文化的側面以外に、実際の対面と比較して①通話相手の実在感（存在感）を感じない、②スムーズな意思疎通が難しい、という2点が挙げられる。前者は映像の精細さや立体感が欠如している事が [1]、後者は特にアイコンタクトの欠如（目が合わない事）が大きく影響していると言われている [2]。

今後、5Gによってネットワークの高速化は進み、高画質の映像がやり取りされ、それに伴い高い対面感をもったビデオ通話 [3] が、実際の商用サービスとして実現されることが期待される。一方で、商用サービスとしての技術的実現性やコストを考慮す

ると、前述したビデオ通話の欠けている要素をすべて完璧に満たすのではなく、実用上、十分な対面感を実現し、かつリーズナブルなシステムにすることが求められる。そこでドコモは、ビデオ通話映像の各パラメータが対面感向上にどのように寄与するかを明らかにし、特に大きな寄与を示した要素を満たすシステムを検討、開発した。

本稿では、システム構築に先立ち実施した評価実験と、その結果に基づき開発したビデオ通話システムの詳細、コミュニケーション活性化を促進させるための拡張機能について解説する。

2. 評価実験

システム構築に先立ち、ビデオ通話映像の各パラメータが対面感に対してそれぞれの程度の寄与を示すかを評価し、重点的に押さえるべきパラメータを明確化した。

2.1 評価手順

本評価では対面感を「普段コミュニケーションをとっている家族や友人があたかも同じ場において会話をしている感じ」と定義し、評価用映像から受ける対面感を評価した。評価は、評価結果を安定させる目的で設けた基準条件（全パターン中で最も対面感が高いと想定される映像）と、各パラメータを変動させた評価条件との比較評価とし、9段階のリッカート尺度*1を用いたMOS (Mean Opinion Score)*2評

価とした。評価者は18～30歳未満の一般人で男女各20名、評価時の視聴距離は1.5mで固定である。変動パラメータは従来研究 [1] [2] で対面感向上の効果が認められている、表1に示す4パラメータとした。評価者は、基準条件から単独でパラメータ変更したパターンと、2パラメータを複合的に変更したパターンの計56パターン×3名の俳優役＝総合計168パターンの映像を見た上で、基準映像と比較した対面感を評価した。実際の評価の様子を写真1に示す。

2.2 評価結果

対面感に対する各パラメータの寄与率を導出するために重回帰分析*3を行った。結果、図1(a)に示すように、決定係数*4 R^2 が0.86 (≥ 0.8) となり、実際の評価値と高い相関をもつ重回帰式（モデル式）で対面感を推定できることを確認した。この重回帰式から、各パラメータの対面感に対する寄与率を導出すると、人物表示スケール（33.0%）>目線ブレ幅（26.0%）≒投影画面サイズ（25.7%）>解像度（15.3%）となった。加えて、サービスとしての受容性を測る目的で、「普段コミュニケーションをとっている家族間や友人間で会話する場合に、ツールとして使用したい／使用したくない」という二者択一の評価を同時に行ったところ、図1(b)に示すように、決定係数 R^2 は0.82 (≥ 0.8) となり、その寄与率は目線ブレ幅（33.4%）>人物表示スケール（30.0%）>投影画面サイズ（19.3%）>解像度（17.3%）の順となった。これらの結果より、投影

表1 評価時の変動パラメータ

パラメータ種別	変動条件
解像度（水平）(pix)	1,920*, 1,440, 1,280, 960, 540, 480
人物表示スケール (%)	100*, 67, 50, 33
投影画面サイズ (inch)	100*, 75, 55, 42, 32
目線ブレ幅 (cm)	0*, 10, 20, 30, 40

※基準条件

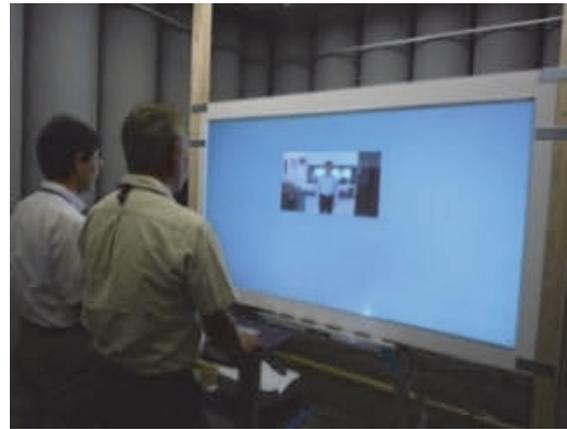
*1 リッカート尺度：アンケートなどで使われる心理検査の回答尺度の一種であり、提示された文に対して回答者がどの程度合意できるかを回答する。一般的に5段階の尺度が用いられるが、7, 9段階の尺度を用いることもある。

*2 MOS：広く用いられる一般的な主観品質尺度の1つ。複数の被験者による主観的な評価を平均した値。

*3 重回帰分析：データ分析手法の1つであり、1つの目的変数を複数の説明変数の線形和で推定しようというもの。この推定式を重回帰式（モデル式）と呼ぶ。



(a)基準条件



(b)人物表示スケール：33%，投影サイズ：32inch

写真1 評価の様子

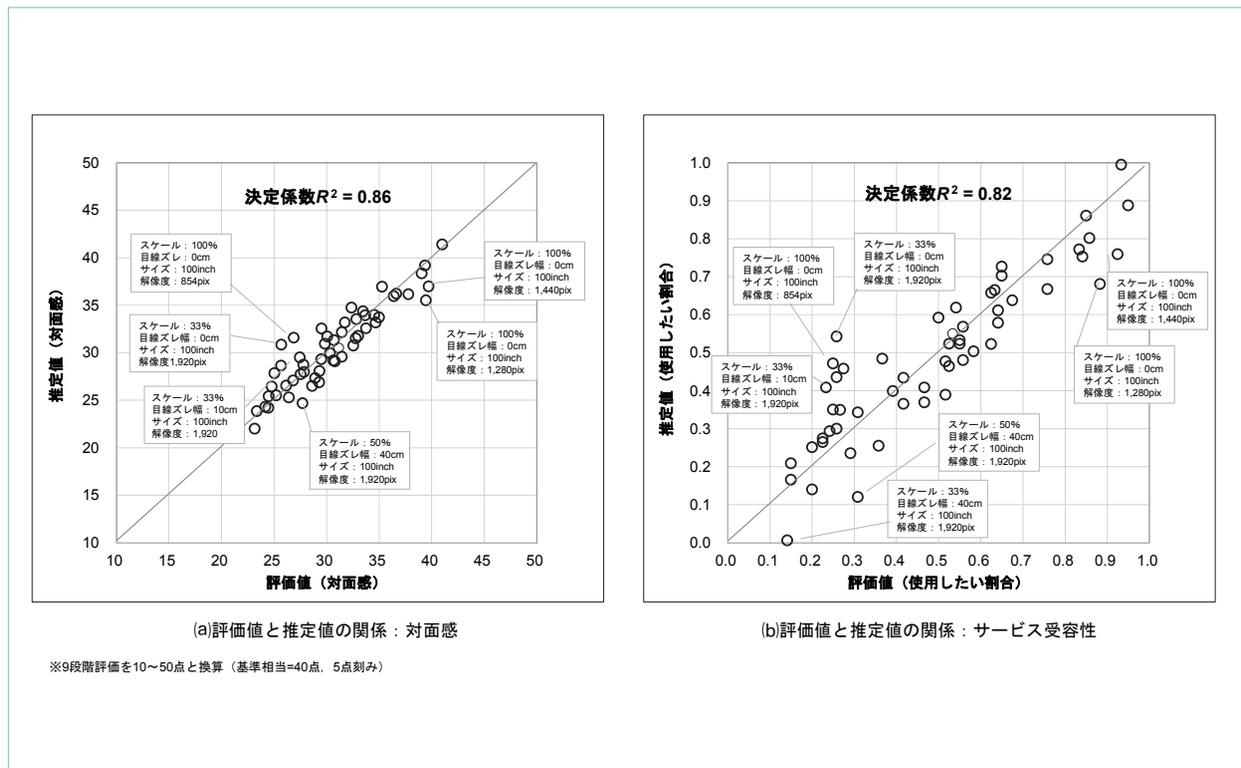


図1 対面感とサービス受容性のグラフ

画面サイズや解像度を大きくする事と比較して、目線ズレ幅を小さくすること（つまり、アイコンタクトを実現すること）、および、等身大に近い人物表

示スケールを実現することが、ビデオ通話システムにおける対面感向上やサービス受容性に高い効果があることが確認できた。

*4 決定係数：重回帰式において推定された値と、実際の評価値との相関度合いを示す指標。一般的に決定係数が0.8以上であれば、推定値と評価値の相関が高い（=重回帰式によって、高い精度で評価値を推定できる）と言うことができる。

3. 正面撮影可能なビデオ通話システム

等身大スケールでの表示については、表示映像からリアルタイムで人物を抽出し、人物に動きがあった場合でも等身大スケールを維持するように映像を拡大・縮小する手法 [4] もある。しかし本システムではその手法は採用せず、一定距離で双方が利用することを想定し、映像を表示するシステムの画面サイズを考慮した最適画角のカメラを用意し、等身大スケールでの表示を可能とした。一方で、技術的にはアイコンタクトを実現するための正面撮影技術（通話相手の映像を画面に表示させつつ、その画面を見ているユーザを正面から撮影する技術）をプロジェクション方式、ディスプレイ方式それぞれで検討し、システムを構築した。

なお、ビデオ通話を実現するソフトウェアとしては、ブラウザ上でのビデオ通話を実現するWebRTC（Web Real-Time Communication）を用いた。

3.1 プロジェクション方式

映像表示部としてプロジェクタとスクリーンを用いたプロジェクション方式は、画面大型化が容易であるというメリットがある。プロジェクション方式で正面撮影を実現する手法としては、液晶スクリーンを用いた時分割方式*5がある [5]。時分割方式を汎用的なデバイスで実現するために、3D表示対応のプロジェクタ、電氣的に透明／不透明の切替えが可能な調光ガラス、シャッタータイミングを外から制御可能なカメラを用いてシステムを構築した。また、調光ガラスをスクリーンとして用いるには映像の輝度・精細さの面で背面からの投影が望ましいため、システム全体の奥行きや設置容易性を考慮して超短焦点のプロジェクタを利用した。

正面撮影の仕組みを図2に示す。3D表示対応のプロジェクタは、サイドバイサイド（SbS：Side by Side）*6方式で入力された60fpsのSbS映像を120fpsの通常映像として投影する。人間の視覚は50Hz程度以上の点滅は残像効果によって常時点灯と視認す

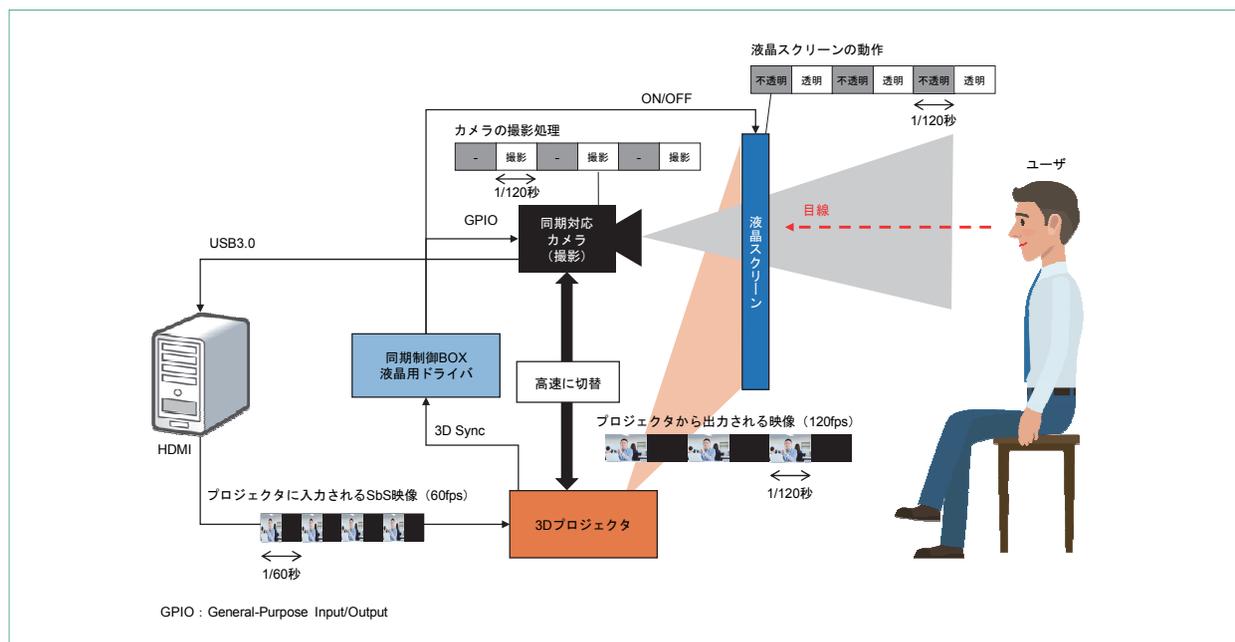


図2 時分割による正面撮影の仕組み

*5 時分割方式：投影-撮影や、3D映像における左目用画像の投影-右目用画像の投影を、時間軸で分割して両立する手法。3D映像における投影では、他に偏向方向で空間的に分割する方式や、波長で分割する方式などがある。

*6 サイドバイサイド (SbS)：異なる2枚の映像に対し、映像の横幅だけをオリジナル映像の半分の解像度として、左右に並べることで、1フレームの映像の中に、解像度が半分となった2枚分の映像を含める方法。主に、3D表示において左目用映像 (L)、右目用映像 (R) を1フレームに収める場合に利用される。

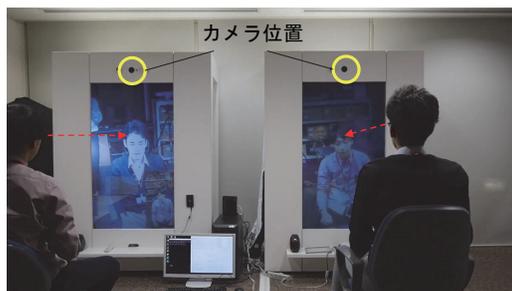
る。そこで、本システムではプロジェクタに入力する映像として通話映像とブランク（黒画面）を並べたSbS映像を用意して3D表示モードで投影することで、疑似的に60Hzの点滅を実現した。3D表示モードでは複数プロジェクタの同期をとるための信号（3D Sync）が秒間60回出力されるため、この信号をカメラとスクリーンに入力することで、プロジェクタで映像を投影しているタイミングにスクリーンを不透明とし、映像を投影していない（＝黒画面を投影している）タイミングでスクリーンを透明としてカメラ撮影を行う。この時分割処理により、ユーザは点滅を視認すること無くスクリーンに投影された映像を見ることができると同時に、スクリーン背面に設置したカメラはユーザを正面撮影することができる。本システムを用いて実際にビデオ通話をしている様子を写真2に示す。本システムによって、等身大に近いスケールで表示された相手の映像を見ながら話すことが実現され、外周カメラで撮影を行う既存方式に比べ、視線が合う自然な会話が可能となることが確認できた。

3.2 ディスプレイ方式

正面撮影の効果はスクリーン方式によって確認できたが、スクリーン方式は背面から投影するプロジェクタを使う方式であるため、スクリーン背面に一定のスペースが必要となる。また、時分割方式で

は、投影される映像の輝度が時分割処理を行わない通常時に比べて理論的に1/2となる。加えて、スクリーンの透明／不透明切替時の液晶応答や3D表示モード時のプロジェクタのクロストーク^{*7}対策の影響により、実質的な輝度は通常時の1/4程度となってしまう。このため、表示される映像が暗く、実在感の低下につながってしまう課題がある。よって、さらなる省スペースで高輝度の映像表示を可能としつつ、正面撮影が可能なシステムが求められる。これを実現するため、我々は透明有機EL（以下、透明OLED（Organic Light Emitting Diode））ディスプレイを用いたシステムも開発した。

透明OLEDは自発光のディスプレイであり、映像表示（発光）時は非常に高輝度の映像を表示できる一方で、映像非表示（非発光）時は40%程度の高い透過率をもつ。また、今回システムで用いた透明OLEDは発光方向に指向性を持ち、ディスプレイ前面からは180度近い視野角をもつ一方で、ディスプレイ後面（背面）からは表示映像がほぼ見えない。よって、時分割処理を行わずとも、透明OLED背面にカメラを設けることで簡易に正面撮影が可能となる。なお、透明OLEDの透明性はユーザ側（映像を見る側）にとっては不要であるため、カメラ前面以外にはディスプレイ背面を黒マスクで覆うことで、映像表示時のコントラストを上げるだけでなく、システム利用時にカメラの存在感を消すことも可能となる。



(a)外周カメラでの撮影（既存方式）



(b)正面撮影（本システム）

写真2 正面撮影の効果

*7 クロストーク：3D表示においては、左目用映像は左目だけで、右目用映像は右目だけで視認されることが理想であるが、左目用映像が右目で、右目用映像が左目で視認されてしまう事がある。この状態をクロストークと呼び、3D酔いや疲れなどの原因とされている。時分割方式の3D表示プロジェクタでは、3Dメガネの液晶応答速度の低下によるクロストークが発生しないよ

う、左目と右目用の映像切替のタイミングで、ブランク（黒）フレームを入れるなどして対策を行っているため、3D表示時の輝度が2D表示時の1/2よりもさらに低くなる場合がある。

このシステムの概要を図3に、システムを利用している様子を写真3に示す。スクリーン方式に比べ、より省スペースで高輝度の映像表示と正面撮影を実現できている様子が確認できる。

4. コミュニケーション活性化の促進

ビデオ通話システムの目的は、通話相手の実在感

向上やスムーズな意思疎通を可能とすることで、遠隔地間でのコミュニケーション活性化を図ることである。コミュニケーション活性化という点においては、一緒に写真を見るような“共有体験”は重要な一要素である。我々はこの共有体験を手軽に実現するために、スマートフォン用アプリと連携し、自身のスマートフォンに入っている写真をシステム上で共有できる機能を設けた。共有時の表示位置につい

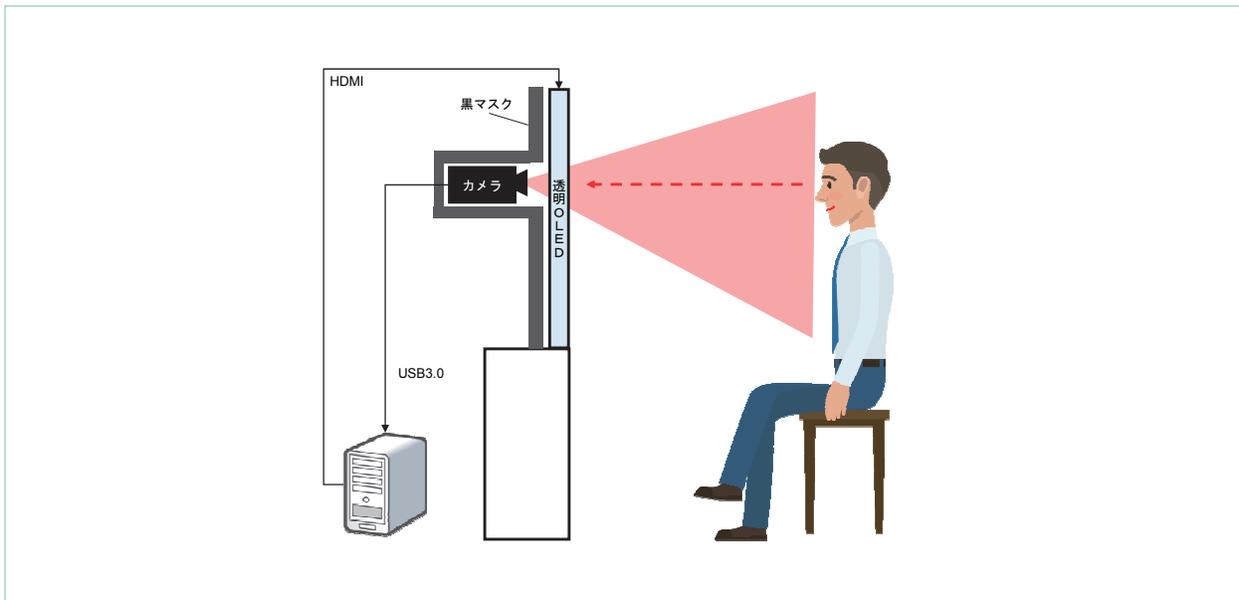


図3 ディスプレイ方式



写真3 ディスプレイ方式のシステムを利用している様子

では、「同じ写真を見ている」感覚を高めるため、鏡面状態（左右対照）で表示した。この様子を写真4に示す。

また、さらなるコミュニケーション活性化に向けては、異言語間でのビデオ通話やテレワーク中に仮想メイクを施す機能 [6] など、多様なユーザーニーズに対応できる機能拡張可能なシステムが求められる。そこで、本システムではビデオ通話で送受信される映像・音声データを別プログラムに渡すインターフェースを用意し、そのデータを使った機能追加が後からできるよう、プラグイン対応としている。写真5に、試作した翻訳プラグインを動作させている様子を示す。本プラグインでは、音声をテキスト化し、それ

を指定した言語に翻訳するAPI（Application Programming Interface）*8を用いることで、異言語間でのビデオ通話であっても、まるで同時通訳が存在するかのように受け手の言語でテロップ表示させることが可能である。

5. あとがき

本稿では、ビデオ通話映像の各パラメータが対面感向上にどのように寄与するかをユーザ評価にて明らかにし、その上で、アイコンタクト実現に必要な正面撮影技術をベースとして構築した「目が合う自然な会話を可能とする対面型ビデオ通話システム



写真4 オブジェクト共有機能



写真5 翻訳プラグインが動作している様子

*8 API：ソフトウェアの機能を他のプログラムから利用できるように切り出したインターフェース。

ム」について解説した。ネットワーク高速化やディスプレイ・カメラなどのデバイス進化がますます進む中で、利用機会がさらに増えることが想定されるビデオ通話において、実用上十分な対面感を満たすシステムの構築をめざした。

ディスプレイ方式の本システムは、2018年12月6、7日に開催された「DOCOMO Open House 2018／ビジネスを、世界を、5Gで革新する。」におけるスマートホームコミュニケーション展示ブース内にてデモを行い、好評を博した。

今後は、パートナーとの実証実験などを踏まえて、商用化に向けて検討していきたい。

文 献

- [1] A. Prussog, L. Mühlbach and M. Böcker : "Telepresence in Videocommunications," Proc. of the Human Factors and Ergonomics Society Annual Meeting, Vol.38, No.3, pp.180-184, 1994.

- [2] L. S. Bohannon, A. M. Herbert, J. B. Pelz and E. M. Rantanen : "Eye contact and video-mediated communication : A review," Displays, Vol.34, No.2, pp.177-185, Apr. 2013.

- [3] KDDI : "Sync Dinner."

<http://connect.kddi.com/sync/dinner/>

- [4] 内田 聡一郎, 足利 えりか, 井元 麻衣子, 我妻 光洋, 日高 浩太 : "イマーシブテレプレゼンス技術 "Kirari!" のコンセプト," 第43回画像電子学会年次大会予稿集, T2-2, 2015.

- [5] H. Ishii and M. Kobayashi : "ClearBoard : A Seamless Medium for Shared Drawing and Conversation with Eye Contact," Proc. of CHI'92, pp.525-532, May 1992.

- [6] 資生堂 : "資生堂, オンライン会議時の自動メイクアプリ「TeleBeauty (テレビューティー)」を開発," Oct. 2016.

<https://www.shiseidogroup.jp/news/detail.html?n=0000000002041>